RTO-MP-10
AC/323(AVT)TP/2

RTO-MP-10

NORTH ATLANTIC TREATY ORGANIZATION

RESEARCH AND TECHNOLOGY ORGANIZATION

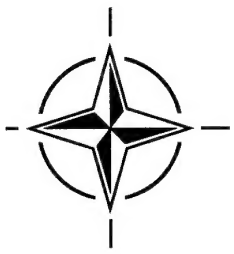
BP 25, 7 RUE ANCELLE, F-92201 NEUILLY-SUR-SEINE CEDEX, FRANCE

RTO MEETING PROCEEDINGS 10

Airframe Inspection Reliability under Field/Depot Conditions

(Degré de fiabilité des visites d'inspection des cellules en dépôt et en conditions opérationnelles)

Papers presented at the Workshop of the RTO Applied Vehicle Technology Panel (AVT) (organised by the former AGARD Structures and Materials Panel) held at Quartier Reine Elisabeth, Etat-Major de la Force Aérienne, Brussels, Belgium, 13-14 May 1998.



Published November 1998

Distribution and Availability on Back Cover

North Atlantic Treaty Organization

Research and Technology Agency

RTA Headquarters: 7, rue Ancelle - 92200 Neuilly-sur-Seine, France

ST/60/4 19 August, 1998

TO: Recipients of RTO Publications

FROM: Scientific Publications Executive

SUBJECT: **RTO Technical Publications**

As you probably know, NATO formed the Research and Technology Organization (RTO) on 1 January 1998, by merging the former AGARD (Advisory Group for Aerospace Research and Development) and DRG (Defence Research Group). There is a brief description of RTO on page ii of this publication.

This new organization will continue to publish high-class technical reports, as did the constituent bodies. There will be five series of publications:

AG **AGARDographs** (**A**dvanced **G**uidance for **A**lliance **R**esearch and **D**evelopment), a successor to the former AGARD AGARDograph series of monographs, and containing material of the same long-lasting value.

MP **Meeting Proceedings**: the papers presented at non-educational meetings at which the attendance is not limited to members of RTO bodies. This will include symposia, specialists' meetings and workshops. Some of these publications will include a Technical Evaluation Report of the meeting and edited transcripts of any discussions following the presentations.

EN **Educational Notes**: the papers presented at lecture series or courses.

TR **Technical Reports**: other technical publications given a full distribution throughout the NATO nations (within any limitations due to their classification).

TM **Technical Memoranda**: other technical publications not given a full distribution, for example because they are of ephemeral value only or because the results of the study that produced them may be released only to the nations that participated in it.

The first series (AG) will continue numbering from the AGARD series of the same name, although the publications will now relate to all aspects of defence research and technology and not only aerospace as formerly. The other series will start numbering at 1, although (as in the past) the numbers may not appear consecutively because they are generally allocated about a year before the publication is expected.

All publications, like this one, will also have an 'AC/323' number printed on the cover. This is mainly for use by the NATO authorities.

Please write to me (do not telephone) if you want any further information.

G.W.Hart

RTO-MP-10
AC/323(AVT)TP/2

NORTH ATLANTIC TREATY ORGANIZATION

RESEARCH AND TECHNOLOGY ORGANIZATION

BP 25, 7 RUE ANCELLE, F-92201 NEUILLY-SUR-SEINE CEDEX, FRANCE

RTO MEETING PROCEEDINGS 10

Airframe Inspection Reliability under Field/Depot Conditions

(Degré de fiabilité des visites d'inspection des cellules en dépôt et en conditions opérationnelles)

Papers presented at the Workshop of the RTO Applied Vehicle Technology Panel (AVT) (organised by the former AGARD Structures and Materials Panel) held at Quartier Reine Elisabeth, Etat-Major de la Force Aérienne, Brussels, Belgium, 13-14 May 1998.

The Research and Technology Organization (RTO) of NATO

RTO is the single focus in NATO for Defence Research and Technology activities. Its mission is to conduct and promote cooperative research and information exchange. The objective is to support the development and effective use of national defence research and technology and to meet the military needs of the Alliance, to maintain a technological lead, and to provide advice to NATO and national decision makers. The RTO performs its mission with the support of an extensive network of national experts. It also ensures effective coordination with other NATO bodies involved in R&T activities.

RTO reports both to the Military Committee of NATO and to the Conference of National Armament Directors. It comprises a Research and Technology Board (RTB) as the highest level of national representation and the Research and Technology Agency (RTA), a dedicated staff with its headquarters in Neuilly, near Paris, France. In order to facilitate contacts with the military users and other NATO activities, a small part of the RTA staff is located in NATO Headquarters in Brussels. The Brussels staff also coordinates RTO's cooperation with nations in Middle and Eastern Europe, to which RTO attaches particular importance especially as working together in the field of research is one of the more promising areas of initial cooperation.

The total spectrum of R&T activities is covered by 6 Panels, dealing with:

- SAS Studies, Analysis and Simulation
- SCI Systems Concepts and Integration
- SET Sensors and Electronics Technology
- IST Information Systems Technology
- AVT Applied Vehicle Technology
- HFM Human Factors and Medicine

These Panels are made up of national representatives as well as generally recognised 'world class' scientists. The Panels also provide a communication link to military users and other NATO bodies. RTO's scientific and technological work is carried out by Technical Teams, created for specific activities and with a specific duration. Such Technical Teams can organise workshops, symposia, field trials, lecture series and training courses. An important function of these Technical Teams is to ensure the continuity of the expert networks.

RTO builds upon earlier cooperation in defence research and technology as set-up under the Advisory Group for Aerospace Research and Development (AGARD) and the Defence Research Group (DRG). AGARD and the DRG share common roots in that they were both established at the initiative of Dr Theodore von Kármán, a leading aerospace scientist, who early on recognised the importance of scientific support for the Allied Armed Forces. RTO is capitalising on these common roots in order to provide the Alliance and the NATO nations with a strong scientific and technological basis that will guarantee a solid base for the future.

Printed on recycled paper

Published November 1998

ISBN 92-837-1002-9

Printed by Canada Communication Group Inc.
(A St. Joseph Corporation Company)
45 Sacré-Cœur Blvd., Hull (Québec), Canada K1A 0S7

Airframe Inspection Reliability under Field/Depot Conditions

(RTO MP-10)

Executive Summary

Non-Destructive Inspection (NDI) reliability is the corner stone of the safety-by-inspection approach for continuing airworthiness of aging aircraft and of the damage tolerance philosophy adopted by many of the NATO members as the basis for ensuring continued airworthiness. Inspection reliability data, usually in the form of technique threshold data and Probability of Detection (POD) data are essential for evaluating the applicability of selected inspection techniques. These data also are used to derive inspection thresholds and inspection intervals. Frequency and method of inspection are primary drivers of maintenance costs and therefore there is pressure to delay onset and reduce frequency. Safety depends on inspection reliability; therefore there is pressure to be conservative in defining onset and frequency. These competing aspects can only be properly evaluated with representative inspection reliability data.

The Workshop had the general objective of promoting general discussion on the merits of the whole concept and use of NDI reliability data in the life cycle management process, including both deterministic and probabilistic approaches. The specific aim of the Workshop was to explore the concept of deriving airframe inspection reliability using field inspection results.

Three overview papers were presented from the perspectives of an end user of inspection reliability data, a researcher in the analysis of data to derive reliability information, and an industrial expert in the definition and application of NDI techniques. It was apparent that NDI reliability is a major influence in the definition of techniques to be applied and their frequency. The parameter used to characterize inspection reliability is Probability of Detection (POD) and the generally accepted target reliability is 90 percent POD at a 95% confidence level.

The derivation of POD statistics was explored. Primarily, this is done with "round-robin" evaluation programs. Human factors were identified as a major element that can affect reliability but are not addressed in these evaluation programs. Analytical methodologies used to derive POD statistics from relatively small data sets were presented and it is apparent that the methods are not standard, and this two organizations using the same data could derive different POD values.

Other papers discussed potential ways that field inspection results could be used to derive POD information. Benefits from this approach include the fact that there is a consideration of human factors built into this approach as well as possible cost reductions by avoiding costly round robin programs. Data deficiencies, both in quality and quantity, were cited as an obstacle to progress. Advanced techniques ranging from enhanced optical inspections and radiography through to unique applications of existing eddy current processes were presented from a reliability viewpoint. Automation is a major advance in improving inspection reliability because it reduces human factor influence.

In conclusion, while there was a consensus that inspection reliability information is a fundamental requirement for effective life cycle management, there was no consensus on who "owned" the requirement to develop and validate the data. Is it the regulators, the operators, the NDI development community or the research community? A recommendation arising from the round table discussion was to form a Working Group to define methods to implement NDI reliability assessments from service data.

Degré de fiabilité des visites d'inspection des cellules en dépôt et en conditions opérationnelles

(RTO-MP-10)

Synthèse

La fiabilité des visites d'inspection par des méthodes non-destructives (NDI) est l'une des pierres angulaires des inspections de sécurité pour maintenir l'aptitude au vol des flottes aériennes d'ancienne génération. Elle est aussi l'une des composantes du principe de la tolérance aux dommages subis adopté par bon nombre des pays membres de l'OTAN comme le garant du maintien de la navigabilité.

Les données sur la fiabilité des visites d'inspection, généralement présentées sous forme de renseignements sur les seuils d'identification des caractéristiques techniques et la probabilité de détection (POD), sont indispensables pour l'évaluation de l'applicabilité des techniques d'inspection choisies. Ces données servent également à la détermination des niveaux et des intervalles des visites d'inspection. Les coûts de maintenance dépendent principalement de la périodicité et de la méthode d'inspection adoptées. Par conséquent, il est financièrement intéressant de repousser le début des visites d'inspection et donc une approche conservatrice s'impose en ce qui concerne la définition de la date de début des visites et leur périodicité. L'évaluation de ces aspects divergents ne peut s'obtenir qu'avec des données représentatives de la fiabilité des visites d'inspection.

Cet atelier a eu pour objectif de créer un forum pour faciliter une discussion d'ensemble sur les mérites du concept et l'applicabilité des données de fiabilité NDI pour la gestion du cycle de vie, en tenant compte des approches tant déterministes que probabilistes. L'atelier a eu pour thème particulier l'examen de l'état actuel des connaissances technologiques en ce qui concerne l'obtention de la fiabilité dans le domaine de l'inspection des cellules en conditions opérationnelles.

Trois communications donnant un aperçu général de la question ont été présentées. Elles représentaient trois points de vue différents: celui d'un utilisateur des données sur la fiabilité des visites d'inspection, celui d'un chercheur dans le domaine de l'analyse des données intéressé par l'extraction de données sur la fiabilité et celui d'un représentant de l'industrie, spécialisé dans la définition et l'application des techniques NDI. Il est apparu très clairement que la fiabilité des techniques NDI est un facteur important pour définir les techniques à appliquer ainsi que leur périodicité. Le paramètre utilisé pour caractériser la fiabilité des visites d'inspection est la probabilité de détection (POD) et le degré de fiabilité généralement admis est de 90% du POD pour un coefficient de confiance de 95%. Cet objectif de fiabilité a été discuté dans le détail. De l'avis général, quoique ces valeurs soient normalement appropriées, elles ne peuvent pas être considérées comme définitives.

Plusieurs communications ont concerné l'origine des statistiques sur le POD. Ces statistiques sont généralement obtenues par le biais de programmes d'évaluation "en rond". Les facteurs humains ont été identifiés comme un élément important, pouvant affecter la fiabilité, mais ces programmes d'évaluation n'en tiennent pas compte. Des méthodologies analytiques utilisées pour l'extraction de statistiques POD d'ensembles de données relativement restreints ont été présentées. Il a été constaté que les méthodes ne sont pas uniformisées et que, par conséquent deux organisations travaillant sur les mêmes données pourraient obtenir des résultats différents.

D'autres communications ont examiné les résultats de visites d'inspection réalisées en conditions opérationnelles comme sources de données POD. L'un des avantages de cette approche est qu'elle tient compte des facteurs humains. En plus, elle permet de réduire les coûts en évitant les programmes d'évaluation "en rond" coûteux. Des carences dans les données, tant quantitatives que qualitatives, ont été citées comme un obstacle au progrès.

Des techniques avancées allant des inspections optiques améliorées et de la radiographie jusqu'à certaines applications uniques de méthodes d'inspection par courants de Foucault ont été présentées du point de vue de la fiabilité. L'automatisation représente un pas important vers l'amélioration de la fiabilité des visites d'inspection, car elle réduit l'influence du facteur humain.

En conclusion, bien qu'un consensus se soit dégagé sur le fait que la fiabilité des visites d'inspection est essentielle à la gestion effective du cycle de vie, aucun consensus n'a été trouvé en ce qui concerne la "responsabilité" du développement et de la validation des données. Est-ce que cette tâche incombe aux régulateurs, aux opérateurs, à la communauté de développement du NDI ou bien aux chercheurs? Lors de la table ronde il a été proposé de former un groupe de travail afin de définir des méthodes permettant de mettre en œuvre les évaluations de la fiabilité des techniques NDI obtenues à partir de données opérationnelles.

Contents

SESSION III: ANALYTICAL ISSUES RELATED TO GENERATION AND USE OF PROBABILITY OF DETECTION DATA

SESSION IV: PRACTICAL EXPERIENCE AND CASE STUDIES

Preface

Inspection reliability is one of the corner stones of the safety-by-inspection approach for continuing airworthiness of aging aircraft and of the damage tolerance philosophy adopted by many of the NATO members as the basis for ensuring continued airworthiness. Inspection reliability data, usually in the form of technique threshold data and Probability of Detection (POD) data are essential for evaluating the applicability of selected inspection techniques. These data also are used to derive inspection thresholds and inspection intervals. Frequency and method of inspection are primary drivers of maintenance costs and therefore there is pressure to delay onset and reduce frequency. Safety depends on inspection reliability; therefore there is pressure to be conservative in defining onset and frequency. These competing aspects can only be properly evaluated with representative inspection reliability data.

Most available NDI reliability data results from dedicated round-robin inspection programs, whereby the same samples are inspected by disparate technicians under laboratory type conditions. These data have been frequently challenged on the basis of non-representativeness of the inspection conditions in terms of environment, access and human factors. Analysis of in-service NDI findings can improve our understanding of the reliability of NDI. This greater confidence in NDI reliability would allow more effective use of NDI for maintaining airworthiness. As an added benefit, by using field data, costs of generating POD statistics could also be reduced.

Significant numbers of in-service detections are occurring, but at present there is no organized process whereby these data are collected and collated for NDI reliability studies. One of the prime purposes of this Workshop and of these proceedings is to raise the profile of using field/depot data for POD determination and to open discussion on the processes under which this data could be collected and analyzed. This intent has been met.

The Workshop was well attended with over 50 attendees. The meeting concluded with a well attended Round Table Discussion. A summary of the main issues and recommendations arising from the presentations and discussions is provided in the Recorder's Report by Professor Doctor J. Schijve.

D.L. Simpson
Chairman
Sub-Committee on
Airframe Inspection Reliability

SUB-COMMITTEE MEMBERS

Chairman

Mr D. SIMPSON
Chief, Structures
Structures & Materials Laboratory
Institute for Aerospace Research
NRC
Montreal Road
OTTAWA, Ontario K1A 0R6

H. Goncalo	-	PO	S. Sampath	-	US
G. Günther	-	GE	R. Servent	-	SP
G. Huseby	-	NO	J. Vantomme	-	BE
P. Heuler	-	GE	W. van der Hoeven	-	NE
J.P. Immarigeon	-	CA	J. Waldman	-	US
S. Paipetis	-	GR	S. Welburn	-	UK
A. Salvetti	-	IT			

Panel Executive

Dr. J.M. CARBALLAL, SP

Mail from Europe:

RTA-OTAN/AVT
BP 25
7, rue Ancelle
F-92201 Neuilly-sur-Seine Cedex
France

Mail from US and Canada:

RTA-NATO/AVT
PSC 116
APO AE 09777

Tel. 33 (0)1 55 61 22 90 & 92
Telefax 33 (0)1 55 61 22 99 & 98

Technical Evaluation Report

by

J. Schijve
Professor of Aircraft Materials (emeritus)
Faculty of Aerospace Engineering
Delft University of Technology
Kluyverweg 1, 2629 HS Delft, Netherlands

Introduction

The former AGARD Structures and Materials Panel (now: Applied Vehicle Technology Panel) organized a workshop on *Airframe Inspection Reliability under Field/Depot Conditions*, held on 13-14 May 1998 Brussels with Mr. D. Simpson as the chairman. The program contained 21 papers with a lecture of Dr. J.W. Lincoln as the lead paper.

Although the title of the workshop suggests a well defined and specific topic for discussion, the variety of the papers is large. The aim of the present Technical Evaluation Report is to discuss the coverage of the papers, in order to see what they have in common and where there is a diversity of approaches. Trends and experience presented in the papers are evaluated and some directions for future developments are emphasized Needless to say that the evaluation and the recommendations reflect the opinion of the Reporter.

References to the papers are made by quoting its number in the program of the workshop. Two papers [12,17] of the planned program were not presented nor made available for inclusion in the proceedings. Some comments of authors, not in the papers but given during the presentation, are used in this report.

The variety of papers

The papers came from different countries, viz.: USA 7 papers, UK 4 papers, Canada 3 papers, Germany 2 papers and one paper from Japan, Italy, the Netherlands, Spain and Turkey respectively. In view of the variety of papers some correlation could occur with the affiliation of the parties involved:

- the aircraft operator (military and civil operators)
- the aircraft industry
- research organizations
- airworthiness authorities

Actually, the papers do not easily fit in this framework, but it could be asked if there are different approaches for military and for civil aircraft operators. They have significantly different operational constraints. Civil operators want to fly the aircraft as much as possible. Maintenance and downtime are well recognized burdens. The economically competitive environment is also a highly dominant aspect for the airlines. Nowadays, economics are important for air forces also, but for different reasons. Life extension programs are considered for economic reasons, although they can require extensive and costly inspections.

Civil aircraft operator problems are discussed in [2,7,8,14,15]. Some more typical aspects of inspecting transport aircraft are visual inspections discussed in [15] and fatigue cracks in lap joints in [8].

Two differences between military and civil aircraft operators became apparent. Boeing [7] applies a safety factor of 3 to determine the repeat inspection period. The military regulations, the USAF for example [1], typically require a factor of 2. Secondly, the civil industry does not seem to use POD quantitatively for determining inspection intervals but rather as a convenient way of comparing inspection process performance [2]. Civil regulators, the FAA for example, are becoming more interested in the evaluation and quantification of inspection reliability and POD definition as evidenced by their sponsorship of research programs and NDI evaluation facilities [11].

An obvious variation of paper subjects is related to the type of structure considered.
Three categories are:

1. Al-alloy aircraft structures
2. Engine components (3 papers)
3. Composite structures (2 papers)

Two papers of the second group [6,10] deal with small crack in engine disks, while the third one [20] covers a new X-ray technique for turbine blades. The two papers in the third category are concerned with detection of impact damage [18] and with automatic C-scanning of composite parts [17].

In the larger first category (15 papers) the variety of papers is still significant. Major emphases are on:

a. NDI techniques, mainly eddy current.
b. POD approach and analysis.
c. Missed cracks and false calls.
d. Two types of inspections, and for which defects?
e. Differences between NDI in the laboratory environment and the field/depot environment.
f. Human factors, education and experience.
g. Economics of NDI.

The discussion below follows the above headings.

NDI techniques

The most frequently quoted NDI technique in the papers is the eddy current (EC) technique. This technique has been available for a long time. An early and noteworthy case was the application to find cracks in bolt holes of the single spar Bristol freighter around 1960 after a catastrophic failure had occurred in this non-fail safe structure. The EC inspection is still superior to other NDI techniques for indicating small cracks in the bore of a hole, see [21]. At the same time the EC technique can also be most useful for surface cracks in general. Heida and Grooteman [5] mention a reliably detectable crack size ranging from 2.5 mm (0.10") to 6.4 mm (0.25") for in-service inspections of the F-16. Regler [16] discussed the EC technique used to find cracks at drain holes in ribs of the F-4F stabilizer, for which an ingenious procedure had to be developed in view of the difficult accessibility through the holes of removed fasteners.

A high reliability for finding small cracks is most essential for engine disks inspections [6,10]. As suggested by Keller et al. [6], it can be used for this application adopting the retirement for cause principle. They developed an impressive robotic EC scanning of disks with a computerized output. They found real cracks in the order of 0.5 to 0.7 mm, approximately 1/3 of the critical crack length. A reliably detectable crack depth of 0.127 mm (0.005") is mentioned. In view of this high sensitivity they had to consider the problem of abnormal reject levels and how to avoid it. Even scratches could give defect signals. It illustrates that any application requires a thorough development study for each specific component environment. This is especially true for engine disks, where critical crack lengths are still small, while they should be associated with a dangerous situation for the aircraft.

Forsyth et al. [10] compared the EC technique applied to engine disks with ultrasonic and X-ray inspections. The EC technique clearly outperformed the other two techniques.

For the Al-alloy aircraft structures the situation can be highly different. Sizable cracks do not necessarily imply a risky situation in a damage tolerant aircraft structure. According to Boeing full-scale test results of an old 737, the crack growth life of a lap splice in the fuselage skin from the visual detection threshold to the linkup threshold of cracks could be in the order of 10000 flights. The crack growth life from an initial high frequency EC detection (first layer) to link up could even be in the order of 25000 flights. Apparently, there is ample time for crack detection in this case. However, as pointed out by Lincoln [1], such results depend on the statistics of several structural and operational variables.

The EC technique is also capable to indicate invisible cracks in so-called 2nd -layer situations. That is important for fatigue cracks in lap joints of sheet material (pressurized fuselages, Aloha accident). However, as discussed by Mullis [19], an automated ultrasonic scan method was preferred for finding fatigue cracks in spanwise splices between planks of the wing of the C-141, with relatively large layer thicknesses varying from 7.0 to 21.0 mm (0.275 to 0.825").

Both EC and ultrasonics are sophisticated methods if compared to visual inspections, including penetrants and magnetic methods. Fatigue cracks usually start at some type of a notch. The notches give already response signals different from areas were notches are absent. It then is necessary to discriminate between responses of notches with small cracks and notches without cracks. The problem was evident in the above mentioned C-141 spanwise splice examinations [19]. Experience with the NDI techniques is then essential as well as a good understanding how the structure can produce false calls. Notches are not a problem for visual inspections. It could even be said that they help to focus on locations where cracks may arise. As shown in the paper by Asada et al. [15] the technical conditions for visual inspections can have a significant influence on the success of visual inspections.

A special optical technique is the D-sight system proposed by Komorowski and Gould [8]. The method is capable to measure small out of plane displacements, which can be due to corrosion between the two sheets of a lap joint, see paper [8], or in a composite structure due to impact damage, see paper [18]. Those are situations which can lead to further deterioration of the integrity of an aircraft structure. Also D-sight should be considered to be a sophisticated method which requires an exploratory development program for each particular application and environment. As an example, the reflectivity of the composite panels studied in [18] is insufficient and it requires some surface treatment as part of the inspection procedure.

The POD approach

In several papers one of the aims is to determine a POD curve, and in addition arrive at a certain confidence level for that curve. The size of the crack or damage which will be found with a probability of 90% is generally considered to be the significant point of the POD curve. In order to be sure that the result is conservative, a confidence level of 95% is introduced. From a pure statistical point of view, such efforts are questionable. The main reasons are: (1) The POD approach is valid only if all NDI measurements belong to the same statistical population. That can hardly be true under practical conditions. (2) Secondly, a confidence level can be

calculated only if we know the statistical distribution function involved. Even if both conditions were met, we are still faced with the chosen values of 90% probability and 95% confidence. Both values are arbitrary choices as pointed out by Lincoln [1]. Beverly [4] also questioned whether the 90/95 criterion is realistic. Bruce [3] mentioned that POD data are unfortunately often reduced to quoting the 90/95 crack length only. Rummel [9] explained that some 10 years ago the 90% POD level was chosen because of limited data being available, while that 90% point on the POD curve appeared to indicate a turning point to low POD values for smaller cracks.

It is rarely realized that a higher probability of detection combined with a lower confidence level may very well give the same result. So why do we make the choice that is usually made. There is no statistical criterion for an optimal choice. Should it then be avoided to make this kind of choices? Certainly not, but they must be based on practical considerations, depending on the relevant NDI environment. For instance, are we really satisfied with a 90% probability for detection of a certain crack size? If the number of inspection locations is very large, could we afford to miss 10% of the cracks? Of course the application of a confidence level gives an extra safety margin (which it really is). But if scatter would be low, the margin is small and a substantial number of cracks can be missed. Maybe, that a 95% detection probability should then be preferred. It anyhow should be realized that a rigorous application of the 90%POD/95%confidence limit implies that missed cracks of that size can still occur, especially if there are many cracks.

The above comments might suggest that the POD graphs have a severely limited usefulness, but that is an incorrect conclusion. A POD diagram contains the results of a large number of NDI measurements. They give useful information on the sizes of cracks, which can be found and about the size of cracks that could be missed. It also gives information about possible scatter. That is all invaluable information. The more data in the graph, the better the informative quality of the graph. Even if a person does not believe in any statistics, it can not be denied that such a graph is a kind of a certificate on the performance and the merits of the NDI inspection technique for the relevant NDI environment. According to Rummel much of this type of information has been compiled on a CD for which he can be contacted.

An interesting approach was discussed by Heida and Grooteman [5]. During periodic inspections on the F-16, if cracks were found, the crack length was measured and extrapolated backwards to estimate the crack length, that should have been present at previous inspections, during which it apparently was missed. A hit could thus be associated with misses in previous inspections. A similar approach was discussed in the presentation of paper [6]. More valuable information is thus obtained than with the hits only.

Missed cracks and false calls

It is noteworthy that "missed cracks" and "false calls"are mentioned in several papers, but consequences are not considered in great detail. Of course, the POD approach recognized the occurrence of missed cracks. It is accounted for by safety factors on inspection periods, assuming that a missed crack will be found next time. According to Lincoln [1] the present state of the art has led to a rather low accident rate of aircraft structural failures. However, the situation for engine disks is more delicate.

False calls do not impair the reliability of the aircraft structure, but they are economically undesirable. It should be expected that a hit is followed by an independent second inspection by another inspector to be sure that it is not a false call, but such an advice is not presented in the papers.

In several papers it is said that increasing the sensitivity of inspection techniques is desirable. It will decrease the detectable crack size a_d and it could then lead to longer repeat inspection periods. However, increasing the sensitivity may also lead to more false calls. It then may be questioned whether reducing or eliminating adverse human factor effects could be a better approach.

The situation is different for engine disks, and as pointed out by Lincoln in the discussion, also for several helicopter components. Although those components can still be certified under regular damage tolerance requirements, small fatigue cracks can lead to rather critical situations. A high sensitivity then is highly desirable.

Two types of inspections, and for which defects?

The title of the workshop is Airframe Inspection Reliability under Field/Depot Conditions. The two major keywords are: inspection reliability and field/depot condition. It was pointed out in some papers that two types of inspections can be specified: (i) Regular inspections and (ii) Special inspections.

Regular inspections are made for all kinds of still unknown and not systematically expected defects, especially corrosion and impact damage. Such inspections occur at fixed periods, not primarily depending on fatigue considerations. These inspections are also referred to as unguided inspections. These inspections are mainly done visually. It is true that corrosion and impact damage in general do not have the same impact on the structural integrity as fatigue cracks could have. However, the damage can initiate fatigue cracks later, as was shown in

the past by several catastrophic aircraft accidents. Komorowski et al. [8] pay much attention to corrosion in their paper, in particular to corrosion in fuselage lap joints with an adverse affect on fatigue. D-sight appears to be a promising technique to detect this kind of corrosion. However, a POD analysis to corrosion inspections does not appear to be feasible.

If defects are found in regular inspections it must be decided if the damage should be removed or repaired. Fatigue cracks can also be found in regular inspections, and again it must be decided whether immediate repair is necessary, or whether flying can still be continued and how long. The aircraft integrity with cracks has to be considered, also because similar cracks may occur at similar locations in all aircraft of a fleet.

Special inspections for fatigue cracks or other defects are requested if it is known that they can occur at special locations. It turns out that most papers are considering fatigue cracks only. The eddy current technique and the ultrasonic examination are options if the crack can not be observed visually, or if the crack is still too small to be seen. An interesting case is presented by fatigue cracks in the lap joints of pressurized fuselages. Fatigue cracks in the top row can be inspected visually. Cracks in the bottom row can not. They occur in a second layer where the eddy current technique can be useful.

Differences between laboratory and field environment

Returning to the two keywords inspection reliability and field/depot condition several papers refer to the two hot issues:

- *the difference between the laboratory environment and the field/depot environment, and*
- *the human factor.*

The first topic is not a gray area. It is fully acceptable that a new NDI technique is developed in the laboratory on specimens with artificial defects. If the technique is not successful under those conditions, it is hard to believe that it will work in service. Circumstances in the laboratory evaluations can then be made more and more realistic by using samples with real fatigue cracks. But after all, the validation has to be done under real service conditions, i.e. in the shop under field/depot conditions by inspectors who have to carry out inspections on routine instructions.

The evaluation of an inspection technique is first done on specimens with known and well defined defects. It has been noted in some papers that fatigue cracks can be simulated by notches, but it has also been recognized that the NDI response of such artificial cracks is not necessarily the same as for a real fatigue crack. Forsyth et al. [10] compared three types of cracks: EDM notches, laboratory fatigue cracks, and service fatigue cracks. The laboratory fatigue cracks were nucleated from EDM notches in the bore of undersized holes. The EDM part was then removed by reaming the holes. The latter procedure was also adopted by Mullis [19] in order to obtain realistic test articles for developing an ultrasonic method for crack detection in the spanwise joints of the C-141. Of course service cracks are still the best option, because they may have a characteristic surface roughness and oxidation. Rummel [9] pointed out that well reproducible EDM notches could still be useful for comparative sensitivity studies.

The human factor, education and experience

Beyond any doubt, the human factor is recognized in all papers of the workshop. It is difficult to define precisely various human factors and to quantify these factors, but that does not mean that we can not reduce shortcomings related to human factors. Of course it requires some understanding of human factors. It is not just one single factor. Various characteristics are discussed in several papers. Three groups can be labeled by:

1. Experience and knowledge of the inspector.
2. Psychological aspects
3. Physiological aspects

It is generally recognized that inspectors should have a professional education in NDI inspection techniques, a thorough training and practical service experience during a substantial period. Educations can be done to different levels and be directed to specific NDI techniques, and even to a certain type of aircraft as discussed by Beverly [4]. The aircraft industry also insists on qualified inspectors [2,7,21]. Boeing develops inspection procedures for each specific case to such an extent that also the weakest airline must be able to do the inspection with their own equipment. According to Hagemaier [2], human errors should be overcome by proper training and standardizing inspection procedures. Whether that goal will always be realized at a field/depot level is an other question. It is a matter of the "culture" in the organization of the aircraft operator. They should have a good infra-structure, with easy communication channels and motivated people at all levels. That can be difficult. How to cope with inspector-to-inspector variations, or even operator-to-operator variations? As stressed by Spencer [11], the development of an inspection procedure should therefore be done in cooperation between the laboratory and the field inspection teams.

The psychological and the physiological situation of the inspector in his inspection environment was discussed by Hagemaier [2] and Lock [14]. Sometimes inspections occur under pretty difficult conditions. But even under good conditions, it should not be overlooked that inspections for fatigue cracks can lead to distraction of attention in view of the repetitive nature of that work

[2,4]. It is looking for cracks which in the large majority of cases are not present. It has been suggested that this aspect of the human factor could be eliminated by developing automated inspection techniques, but it has received little attention in this workshop. Mullis [19] describes an automated ultrasonic scanning for cracks at fasteners of a spanwise splices. Valdecantos et al. [17] describe automated C-scanning of composite parts. In some quality control problems in production (composites, laminates, adhesive bonded sheet structures) it is quite obvious that automatic inspections are necessary, but that is not in a field/depot environment. Automatic inspections for fatigue cracks in a complex aircraft structure are not that easily realized.

Economics of NDI

If statistics or risk analysis is to be applied to models on the efficiency of inspection procedures, nobody knows how to account quantitatively for the human factor. In all honesty, everybody knows that it is practically impossible. Lincoln [1] offers rather skeptical views about human factor considerations. Lock [14] as an external observer stresses "the largely unquantifiable nature and therefore the implausibility of using human factors".

In any case of a potential fatigue problem, the method to be chosen will depend on the specific shape and accessibility of the component. Disassembly, which should be minimal [7], may be necessary. That applies e.g. to finding cracks in holes in massive parts. The bolt must be removed. These kind of aspects is important for the cost-effectivity of the inspection procedure. It is pointed out in some papers that trade-offs should then made. However, in many cases, it may be questionable whether unambiguous rational calculations can be made. As said by Spence [13]: "The integration of inspections with routine maintenance service schedules plays a critical role in the optimization process, whereas inspection technique costs play a less significant role", and "History shows that the inspection programmes have been based more on engineering judgement and the experience of expert engineers than on in-depth calculations".

Some recommendations

Airframe inspection reliability will remain a matter of concern in the future, especially so for aircraft with life extension programs, and for aging aircraft in general. Secondly, the reliability of inspections is also very important for components, where failure could cause a serious accident. This applies to engine disks, to massive single-load path structures and several helicopter parts. Good inspections can make such components damage tolerant according to the regulations. But if the component is not really a fail-safe item, finding cracks too late can cause a disaster. Both reliability and a high sensitivity of the inspection technique are then essential.

In view of the previous summary and evaluation some recommendations are presented below:

1. A generally felt weak link is associated with all aspects of human factors. Inspection procedures should not only describe what the inspector must do, but also under which conditions it has to be done. The conditions should not be described in general terms only, but also as requirements for the specific inspection task to be carried out.
2. Inspections have the character of routine activities. The quality of inspections must therefore be maintained by regular checks on the performance of the inspectors. Refresher courses and repeat examinations should also be considered.
3. In view of the routine character of inspections and the repetitive nature of finding no cracks, the motivation of the inspector should by systematically encouraged. A good culture of responsible teamwork on the field/depot level must be pursued.
4. Service experience of the operators concerning detection of cracks and other defects, and techniques used, should be documented in a suitable and reader friendly format. That should be made available to all inspection teams, also on the depot/field level.
5. There is an apparent need to define the processes for collecting, collating and analyzing data from service experience. Strict definitions are required for characterizing the detected cracks. Also, to be useful, technical details of the NDI technique used must be captured. The paper by Asada and Sotozaki [15] gives an example of setting up a database for visual inspections. Similar efforts are required for all NDI techniques. The NATO Research and Technology Organization is uniquely placed to contribute to this area because of its broad membership and access to field inspection results.
6. The above recommendations would benefit from an international cooperation between all operators of the same aircraft or the same type of aircraft.
7. Improved inspection techniques are especially desirable for fatigue critical components, where missed cracks can cause an aircraft accident. Automation of the inspections should then be considered if it eliminates the human factor. In certain cases automation may also be more cost-effective.
8. The reliably detectable crack length (a_d) should be selected by considering consequences for the aircraft safety and the cost-effectivity of the inspections. Its value should not necessarily follow from a 90% POD value.
9. The significance of corrosion damage for fatigue crack initiation and growth should be given due attention in the future. Inspection procedures for this topic must be

considered.

10. Lincoln [1] has made a plea for tear down inspections of older aircraft. This should be strongly supported for several reasons. It can reveal crack locations, but also the shapes and the sizes of cracks. Moreover, parts of the aircraft structure, which are not disassembled and torn down, can be used as test articles for inspection training, and for evaluating inspection procedures and new NDI techniques. Finally, the results of tear down inspections are basic evidence for checking prediction models for crack initiation and for crack growth from the "detectable crack length" ad to the "critical crack length" ac. It is rather optimistic to assume that the crack growth curve for a given load spectrum can accurately be calculated. We still have to live with a limited validity, and thus a limited reliability of our prediction models for realistic aircraft load histories and environments and a complex structure. Full-scale flight-simulation tests can improve the situation, but that is an accelerated test for a single load spectrum. Conservative assumptions are made for various input data, but nevertheless safe and reliable flights remain dependent on reliable airframe inspection at field/depot environments.

11. The Round Table Discussion highlighted the requirement for a coordinated effort for collating and analyzing field inspection data results (and tear-down inspections) into a properly structured database. The present Workshop has contributed to this issue by compiling present ideas and experience. There is still a challenge for the future.

Acknowledgment: Mr. D. Simpson (chairman of the panel) has contributed the present Technical Evaluation Report by some worthwhile comments.

ROLE OF NONDESTRUCTIVE INSPECTIONS IN AIRWORTHINESS ASSURANCE

John W. Lincoln
ASC/EN
2530 Loop Road West
Wright-Patterson Air Force Base, Ohio 45433-7101, USA

SUMMARY

Since the seventies, when United States Air Force and the Federal Aviation Administration adopted damage tolerance, much attention has been focused on the reliability of nondestructive inspections of metallic structures. Although there has been considerable effort expended on analyses and tests for many years, there are still serious concerns about the ability to adequately quantify this reliability. This is true for both the widely used deterministic approach as well as the probabilistic approach. The probabilistic approach, which is currently gaining new interest, is particularly difficult because the complete probability of detection (POD) function must be determined. Much of the concern with inspection reliability is associated with the lack of understanding of the difference between the laboratory environment and the field environment. Another concern is the level of competence of the inspector needed to reflect the detection probability developed for the instrument. It is the purpose of this paper to illustrate the importance of understanding the reliability of the inspection process to continued airworthiness. This will be accomplished primarily through probabilistic methods. The paper will also discuss the use of teardown inspections to enhance the quantification of the inspection reliability. Further, it will discuss some of the current efforts to enhance reliability of the inspection process.

1 INTRODUCTION

The United States Air Force (USAF) Aircraft Structural Integrity Program (ASIP) (Ref 1), when initiated in 1958, adopted the reliability approach called "safe life" to ensure the structural safety of operational aircraft. This approach determined the life of an aircraft in service by dividing the successfully demonstrated capability through full-scale cyclic testing by a factor called the scatter factor. Through the years the "safe life" method was in use, the USAF used a range of numbers for the scatter factor, but they never used a number greater than four. With this approach, the USAF found little need for nondestructive inspections. They found inspections of the fatigue test structure valuable for establishing the "safe life' but little else. It was the intent of the USAF to operate an aircraft for the number of hours corresponding to the "safe life" and then retire it.

The ASIP was dramatically changed in the early seventies when the USAF adopted the damage tolerance approach for maintaining flight safety. This change was made because the "safe life" approach used by the USAF did not achieve the desired structural safety (Ref 2). The first USAF application of damage tolerance was on the F-111. It was the result of an unexpected failure of an F-111 in operational service on 22 December 1969. In response to this accident, the USAF convened their Scientific Advisory Board to determine what should be done to preclude further catastrophic events. This board recommended a cold proof test of the F-111 to limit load to ensure that it could operate safely. The number of hours of safe operation could be determined from knowledge of the fracture toughness at the proof test and operational temperatures, and from the crack growth rates. The USAF repeated the cold proof test at the end of the safe operating interval. This process was successful in that the failures experienced in the proof test likely precluded a repeat of the 1969 incident.

As evidenced by the F-111 experience, the inspection of a structure by proof testing is quite effective. It still is the most reliable inspection process available. As found through approximately a dozen failures of F-111s in cold proof tests, it is not always nondestructive. Further, it does not lend itself to all applications. It was effective on the F-111 because the reduced temperature significantly lowered the fracture toughness of the offending steel components of the structure. However, the inspection intervals for the aluminum components were too short to be useful since aluminum does not significantly change its toughness in response to lowering its temperature.

Another opportunity to use the proof test concept for ensuring safety came in the middle seventies with the B-52D aircraft. In this case, the USAF found the wings to be significantly cracked as a result of Southeast Asia operational usage. The USAF planned to replace the wing structure. However, they wanted to continue operating the aircraft in a training environment until they could schedule them into modification. They could have been inspected nondestructively through eddy current and ultrasonic techniques at considerable expense to the USAF. Since the munitions could be removed from the aircraft for the training mission, the limit load factor could be significantly increased. This fact made the proof test viable for the aluminum airframe. This was accomplished rapidly and economically and provided the USAF with safe aircraft for training until the modification was accomplished.

The use of proof testing was also evaluated in the late seventies for the original C-5A wing. In this case, the reinspection intervals were too short for the method to be economically and logistically viable.

In addition to establishing safe operating intervals for the F-111, this pioneering effort in damage tolerance was important for two other reasons. The first is that the approach used on the F-111 was deterministic. Although some aspects such as the fracture toughness and the crack growth rate data had probabilistic considerations, the proof test inspection intervals were deterministically calculated. A second reason for the importance of the F-111 work is that the residual strength was based on limit load. This is a deterministic concept and has been the basis for damage tolerance assessments for both military and commercial aircraft.

2 INSPECTION RELIABILITY

When the USAF adopted damage tolerance, it was apparent that they must make nondestructive inspections an integral part of the process. This was evident in the drafting of the first specification for damage tolerance (Ref 3). In this specification the USAF supposed that an inspection interval of one-quarter of the design life of the aircraft would be acceptable to the logistics community. There was

Paper presented at the RTO AVT Workshop on "Airframe Inspection Reliability under Field/Depot Conditions", held in Brussels, Belgium, 13-14 May 1998, and published in RTO MP-10.

location, such that the joint probability density function for crack length and stress is the product of the respective marginal density functions. The procedure supposes that the crack growth function is not a random number set and it depends only on flight time. Further, the critical stress function (i.e., the stress at which a crack will propagate by rapid fracture) is not a random number set, and it depends only on stress. The procedure may be generalized such that the crack growth and the critical stress functions are only known through their probability distributions.

For each location in the structure that has an effect on the probability of failure, the following functions must be determined:

Probability distribution for crack length at a reference time

Probability distribution for stress

The critical stress function

Crack growth function

Probability of detection function for the inspections

The examples are based on a hypothetical aircraft that has been designed for a life of 30,000 hours. It is assumed that the aircraft is to fly only one mission type. It is supposed that each mission is two hours in length. It is further supposed that there is only one area of the structure required for this calculation and the only significant contribution to the risk from that area is 500 fastener holes. It will also be assumed that for each of these holes, the initial crack distribution may be represented by a crack distribution function developed for the A-7D damage tolerance assessment. Figure 3 shows the crack probability distribution function, and Figure 4 shows the corresponding probability crack density function. For the intact structure, the stress exceedance function for each of these holes is shown in Figure 5. The corresponding stress probability distribution function and stress density functions, derived from the exceedance function, are shown in Figures 6 and 7. Figure 8 shows the residual stress function. The procedure used supposes that the fracture toughness is not a random number set. The crack growth function that the procedure uses to modify the initial crack probability distribution function so that the crack probability distribution has the correct time dependence is shown in Figure 9. The final function that is needed for the calculation of risk is the inspection of inspection function shown in Figure 10.

The single flight probability of failure for the intact structure without inspections is shown in Figure 11. From this figure, it is seen that the risk exceeds the one in 10 million threshold of acceptability (Ref 9) at about 22,000 flight hours. On the basis of the crack growth function shown in Figure 7, the residual stress function shown in Figure 8, and the inspection probability of detection function shown in Figure 10, the damage tolerance inspections may be determined. Based on limit load, the critical crack length is 8.81 millimeters. The initial crack is assumed to be 1.27 millimeters and the inspectable crack length is based on a 0.90 probability of detection at 2.54 millimeters. The point D of the POD function in Figure 10 is the point (2.54,0.9). The safety limit is the time for the initial flaw (or the nondestructive inspectable flaw) to grow to critical crack length. The USAF divides the safety limit by two for their damage tolerance based inspection program. Consequently, the first inspection is at 7,752 flight hours, and the inspection interval following the first inspection is 5,000 hours. The single flight probability of failure for the intact structure with inspections is shown in Figure 12. It is seen that these inspections are quite effective in reducing the risk of failure, and the risk is contained within acceptable limits to 30,000 flight hours. It is clear from this figure that on the basis of the inspection capability assumed and the inspection interval derived from the damage tolerance methodology, the risk is increasing significantly. Therefore, a reduction of the inspection period must be made if it is intended to fly the aircraft significantly beyond its originally intended life of 30,000 flight hours.

The current USAF policy is to recommend inspections at one-half of the safety limit. However, they do not mandate the accomplishment of the inspection until the aircraft reaches the safety limit. The risk assessment procedure permits one to determine the effect of this policy on structural safety. Figure 13 shows the single flight probability of failure for the case where the inspections are made at the safety limit. One sees that the risk is less than that determined for the case of no inspections shown in Figure 11. However, it is seen that there is a significant degradation in risk as compared to inspecting at one-half of the safety limit.

It was noted above that the deterministic damage tolerance method requires knowledge about one point of the probability detection function. The probability of detection function could, of course, be altered significantly and still maintain this single point fixed. A simple example indicates that the effect on risk from minor modifications may not be too significant. One modification chosen for probability of detection function was to suppose that it remained the same except no point of the function exceeded 0.9. Thus, the ordinate of each point of the POD function whose x-projection is equal to or greater than 2.54 is 0.9. This scenario is unlikely to occur in practice. Figure 14 shows the result of this modification. One sees that there is only a slight increase in risk over that shown in Figure 12.

Alternately, one may select a modification to the POD function shown in Figure 10 by assuming that the ordinate of the POD function is zero for each point of the POD function whose x-projection is less than 2.54. The results of this modification are shown in Figure 15. For this example, the risk again is only slightly greater than that shown in Figure 12.

As a final example one may make a radical modification to the POD function by supposing that the ordinate is zero for each point whose x-projection is less than 2.54, and the ordinate is 0.9 for each point whose x-projection is equal to or greater than 2.54. Figure 16 shows the results of using this probability of detection. It may be seen that the risk has increased significantly. However, for this example, the risk is still acceptable if the aircraft retirement time is 30,000 flight hours.

5. CONCLUSIONS

The teardown inspection is by far the most valuable tool to quantify the reliability of a given inspection method. Therefore, one should not ignore such an opportunity since they occur only rarely. It is also valuable to compare results from a teardown inspection with those derived from more conventional methods. This will provide a better understanding of the capability of conventional techniques.

The examples shown serve to illustrate how important inspections may be in maintaining the structural integrity of an aircraft. The damage tolerance derived inspections were able to control the risk acceptably. For the example aircraft chosen, relatively minor changes to the probability of detection function did not significantly affect the probability of failure. However, the examples did show a significant change when one omits the factor of two in determining the inspection intervals. The results shown are for a certain set of assumptions on stress, crack growth rates, and initial crack size distribution. The approach; however, appears to have some merit for assessing the validity of the deterministic damage tolerance approach. Therefore, one should attempt to generate the data needed for the probabilistic calculations.

There is still much to be done to understand the human factor influence on the reliability of the inspection process. Emphasis on automation of the inspection process will serve to eliminate this unknown element in the inspection equation.

6 REFERENCES

1. Department of the Air Force "Aircraft Structural Integrity Program, General Guidelines For," MIL-HDBK-1530C, 30 August 1996.

2. Lincoln, J. W. "Damage Tolerance - USAF Experience", Proceedings of the 13th Symposium of the International Committee on Aeronautical Fatigue, Pisa, Italy, 1985.

3. Department of the Air Force, 1974 "Airplane Damage Tolerance Requirements," Military Specification MIL-A-83444.

4. Berens, A. P., "NDE Reliability Data Analysis", Metals Handbook, Volume 17, 9th Edition: Nondestructive Evaluation and Quality Control, 1988

5. Elliott, W. R., Lincoln, J. W., and Register, D. C. "Assessment and Terminating Action for Widespread Fatigue Damage in C-141 Wing Fuel Transfer Holes", Proceedings of the 18th Symposium of the International Committee on Aeronautical Fatigue, Melbourne, Australia, 1995.

6. Lewis, W. G., Dodd, B. D., Sproat, W. H., and Hamilton, J. M., "Reliability of Nondestructive Inspections—Final Report," SA-ALC/MME 76-6-38-1, December 1978.

7. Lincoln, J. W. "Aging Aircraft - USAF Experience And Actions," Proceedings of the 19th Symposium of the International Committee on Aeronautical Fatigue, 16th Plantema Memorial Lecture, Edinburgh, Scotland, 1997.

8. Spencer, F. et al, "Reliability Assessment at Airline Inspection Facilities," DOT/FAA/CT-92/12, March 1993

9. Lincoln, J.W., "Risk Assessments - USAF Experience," Proceedings of the International Workshop on Structural Integrity of Aging Airplanes, Atlanta, GA, March 1992.

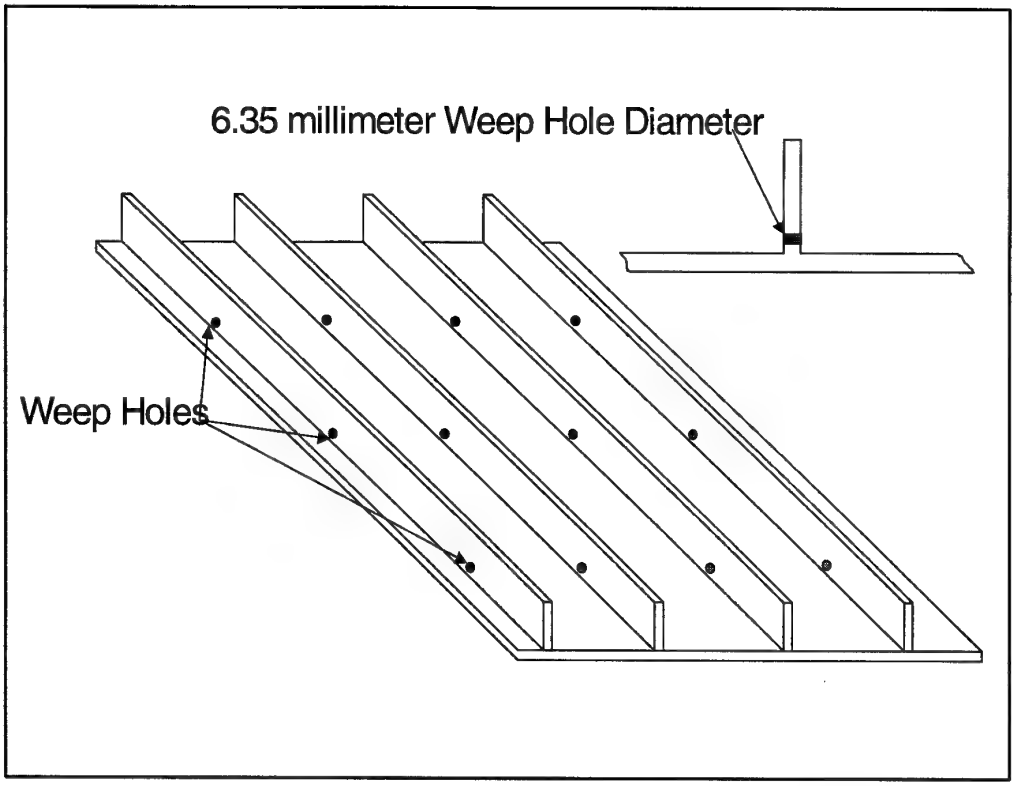


Figure 1 C-141 Inner Wing Lower Surface Panel

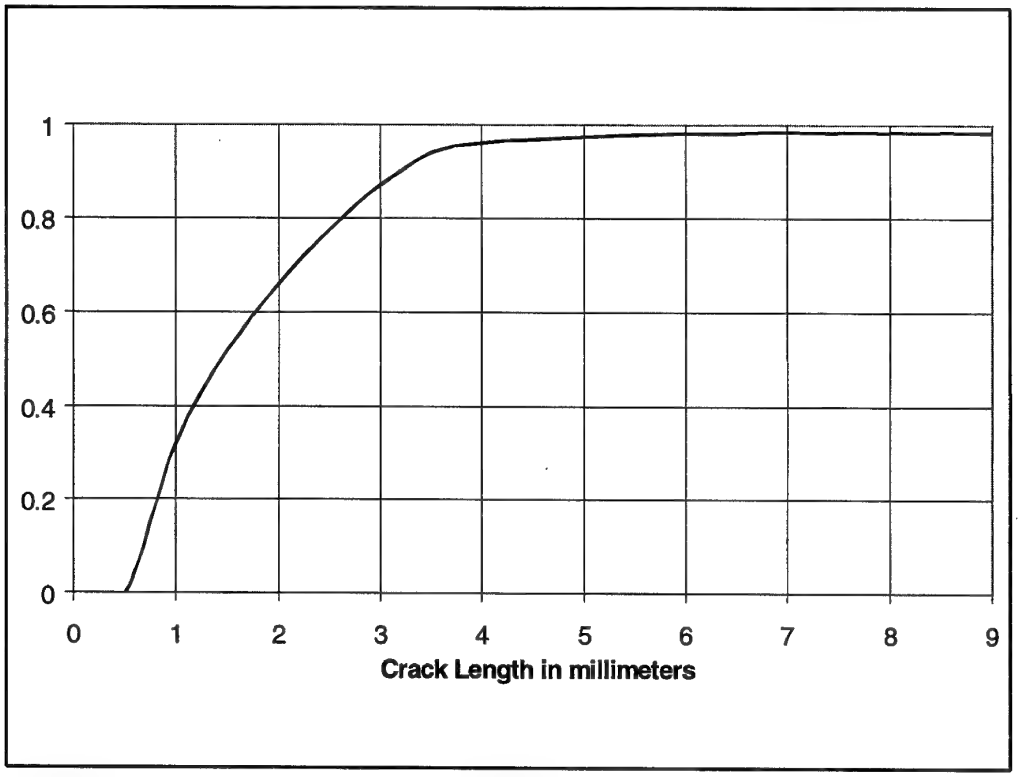


Figure 2 C-141 Wing Teardown Inspection Reliability

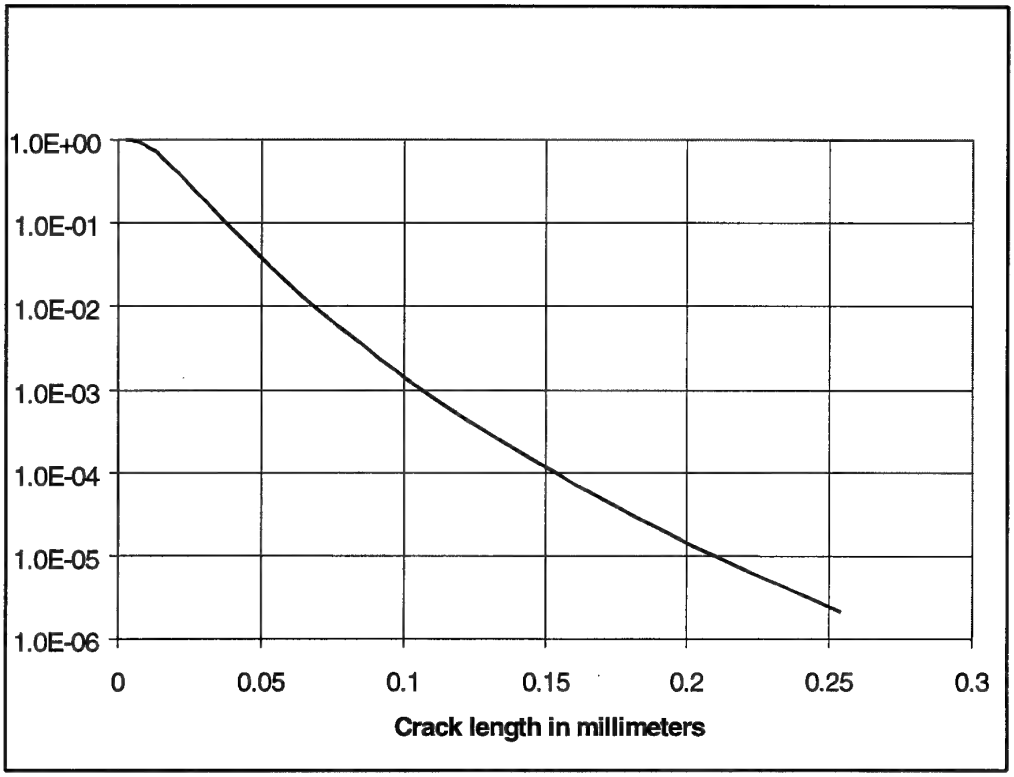


Figure 3 Crack Distribution from A-7D Lower Wing Skin

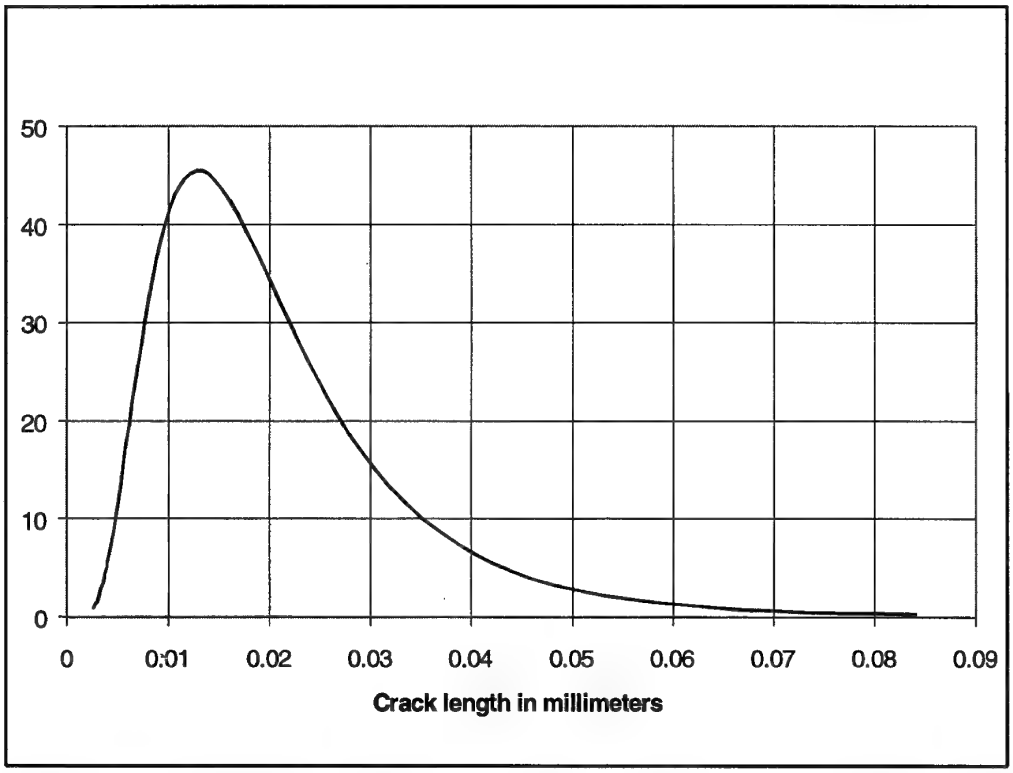


Figure 4 Crack Density Function from A-7D Lower Wing Skin

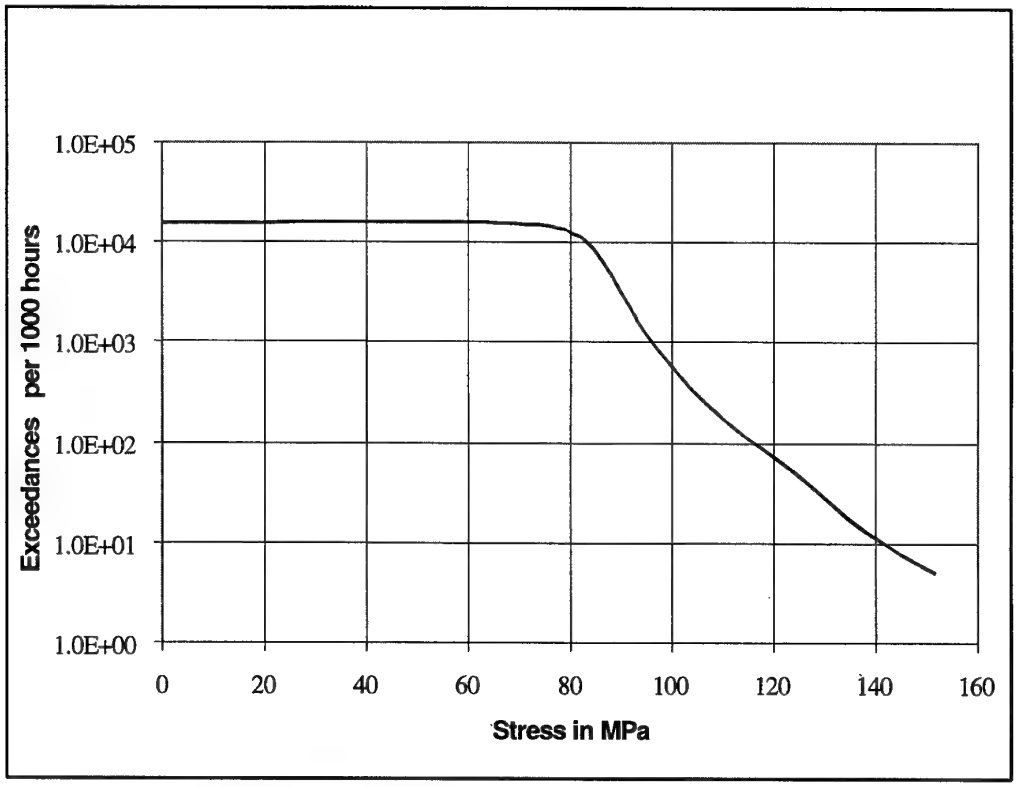


Figure 5 Stress Exceedance Function

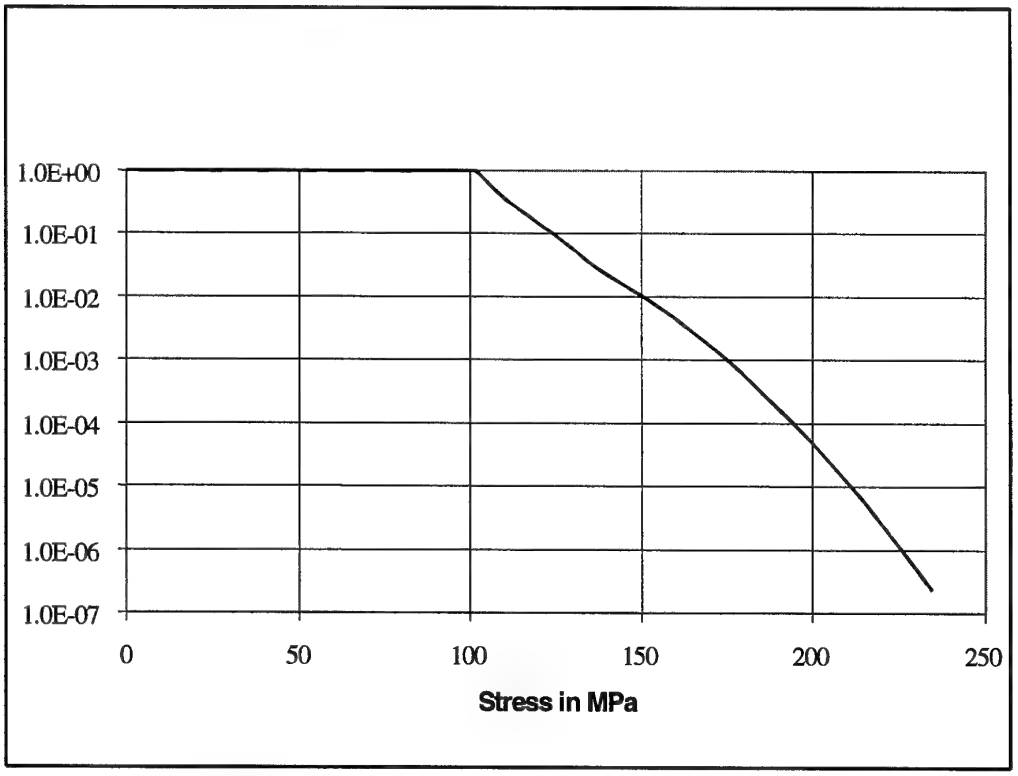


Figure 6 Stress Probability Distribution Function

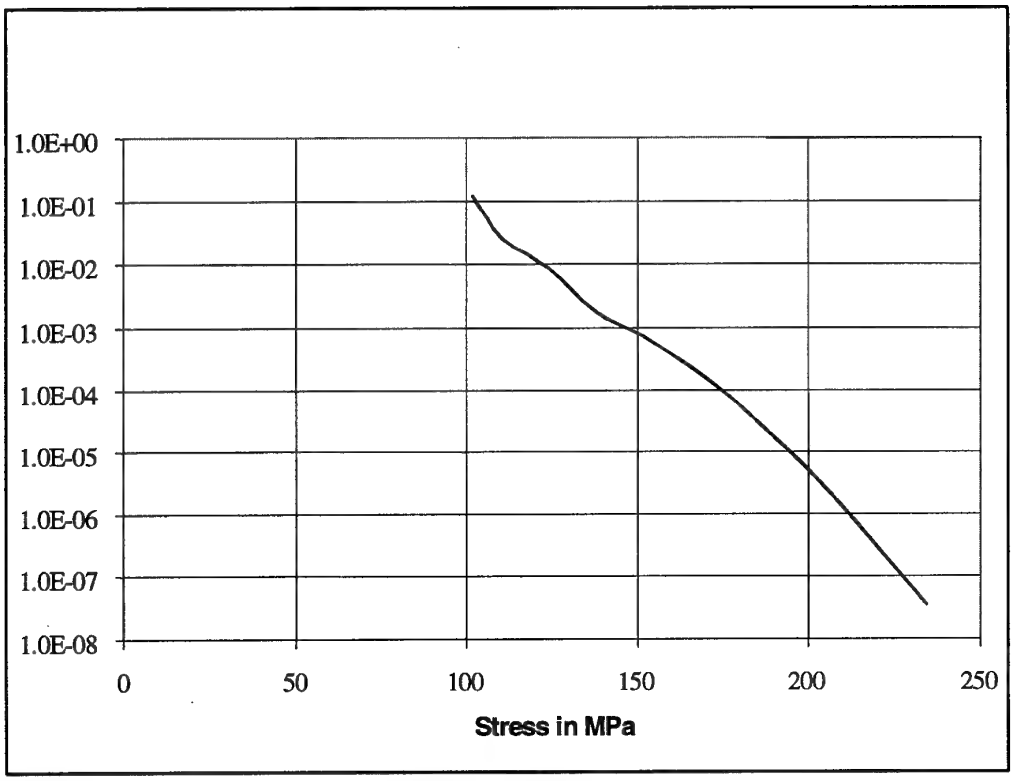


Figure 7 Stress Probability Density Function

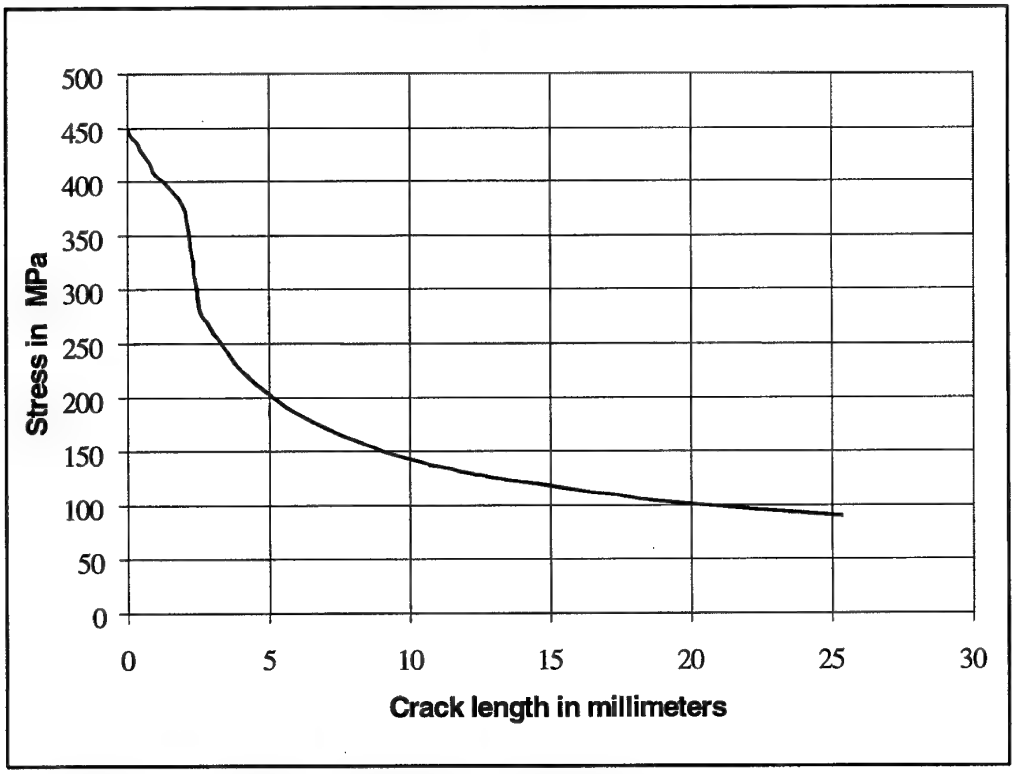


Figure 8 Residual Stress Function

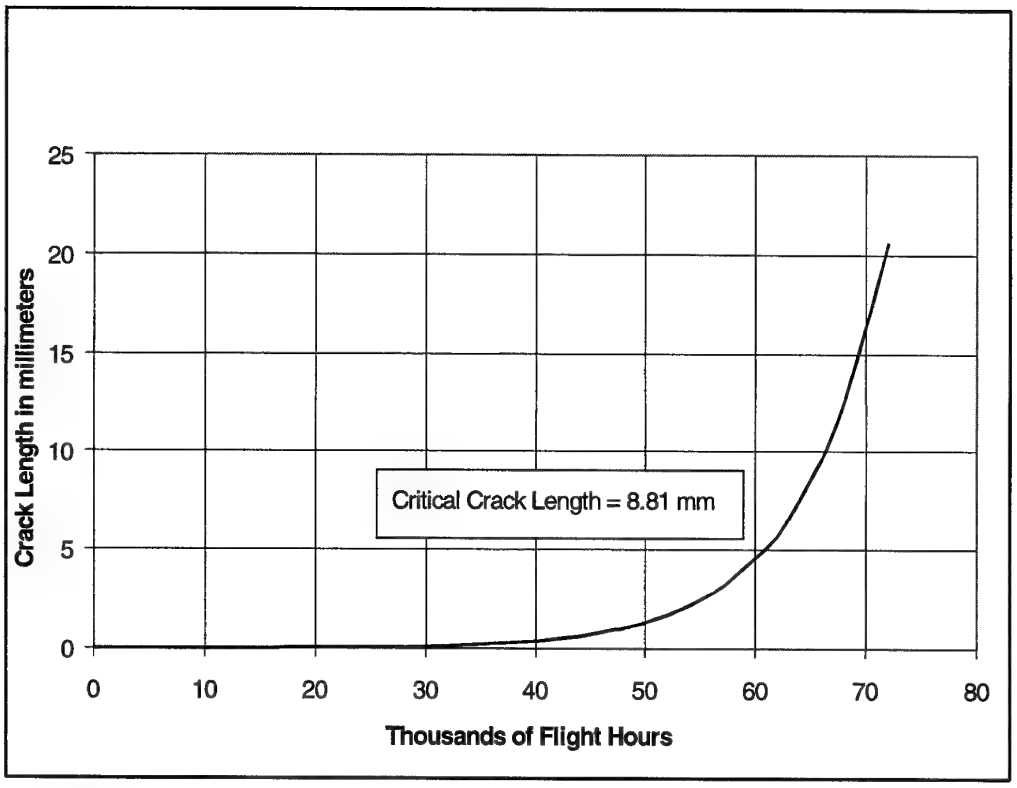


Figure 9 Crack Growth Function

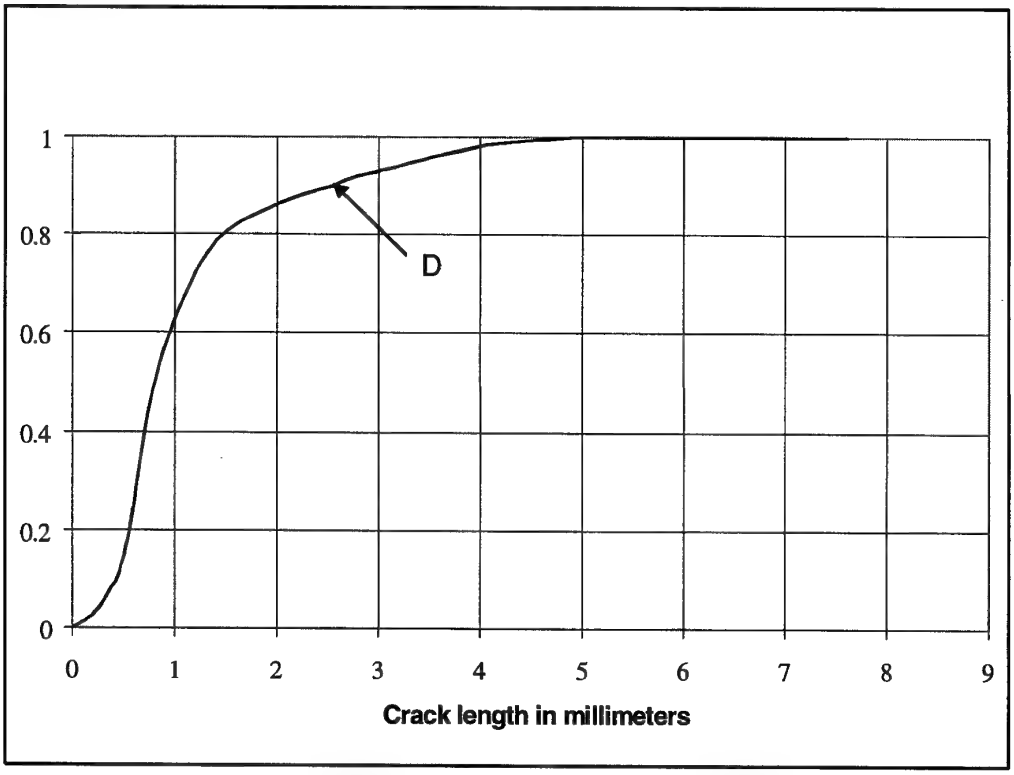


Figure 10 Probability of Detection Function

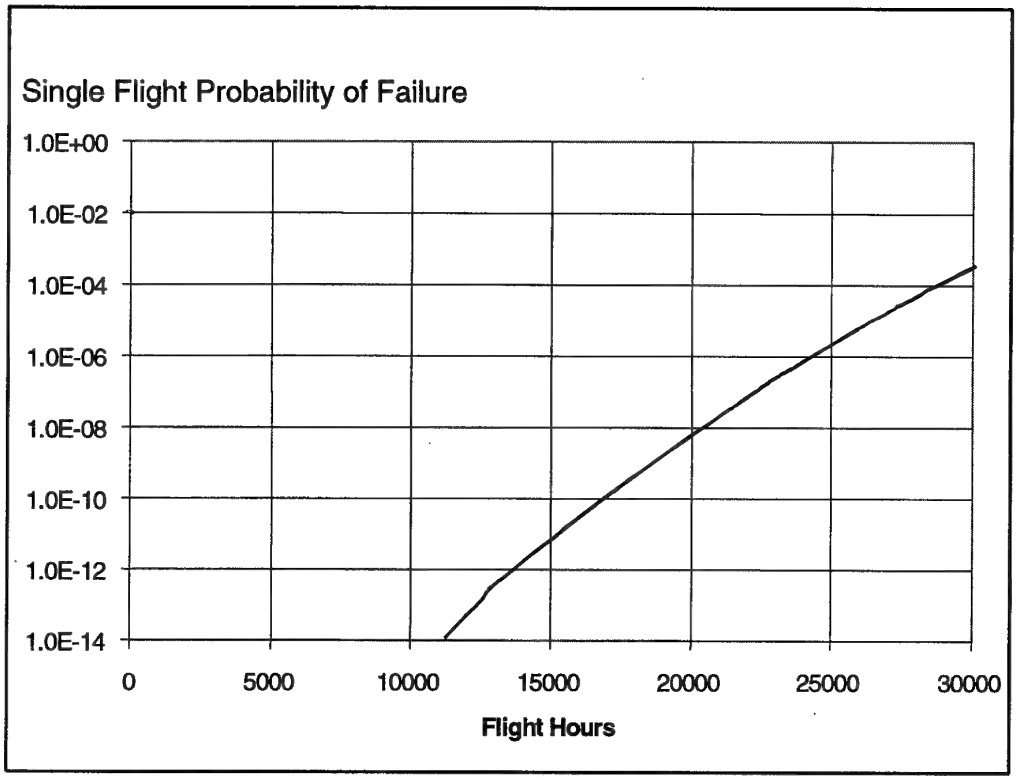


Figure 11 Risk with No Inspections

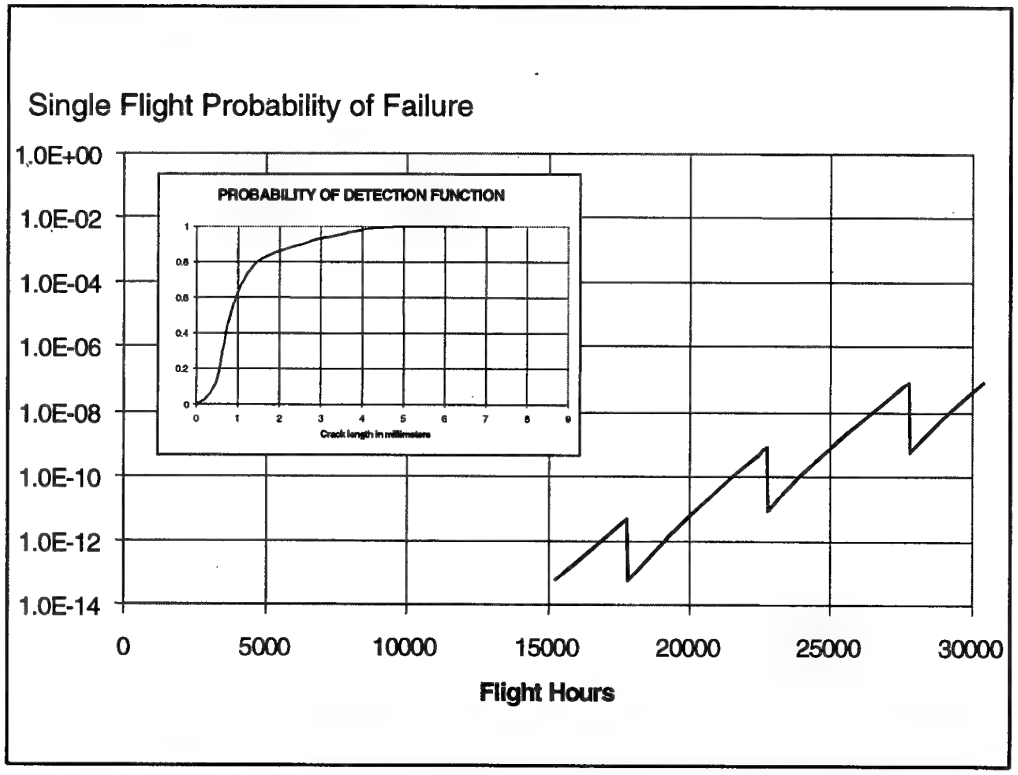


Figure 12 Risk with Damage Tolerance Inspections

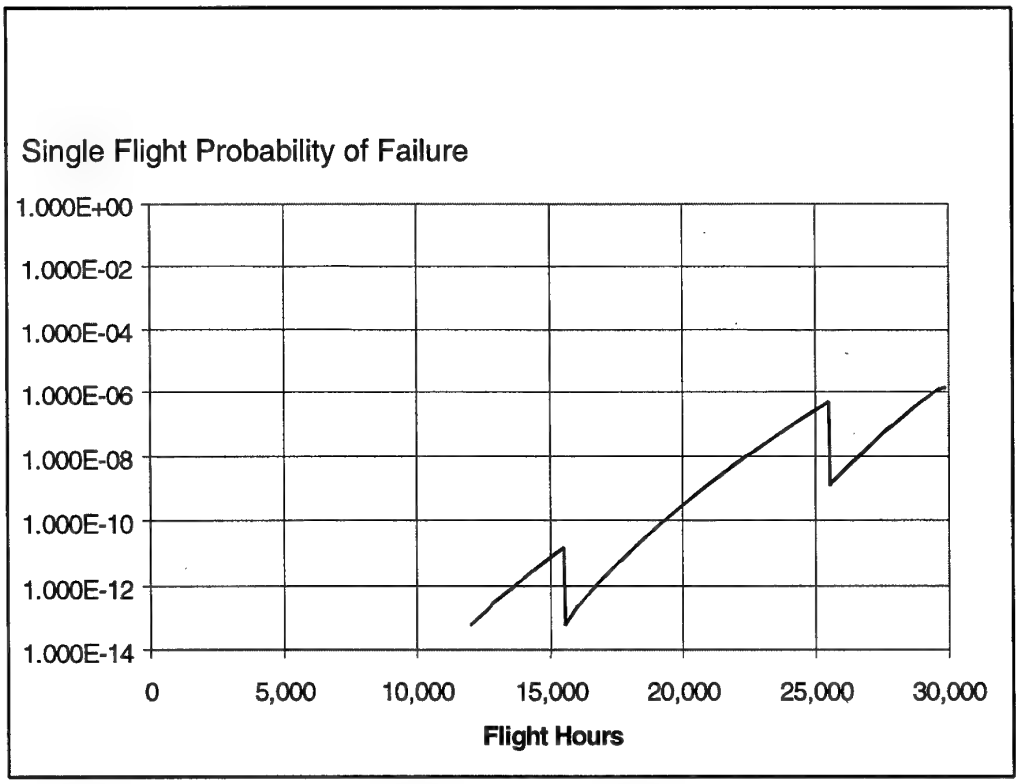


Figure 13 Risk with Reduced Inspections

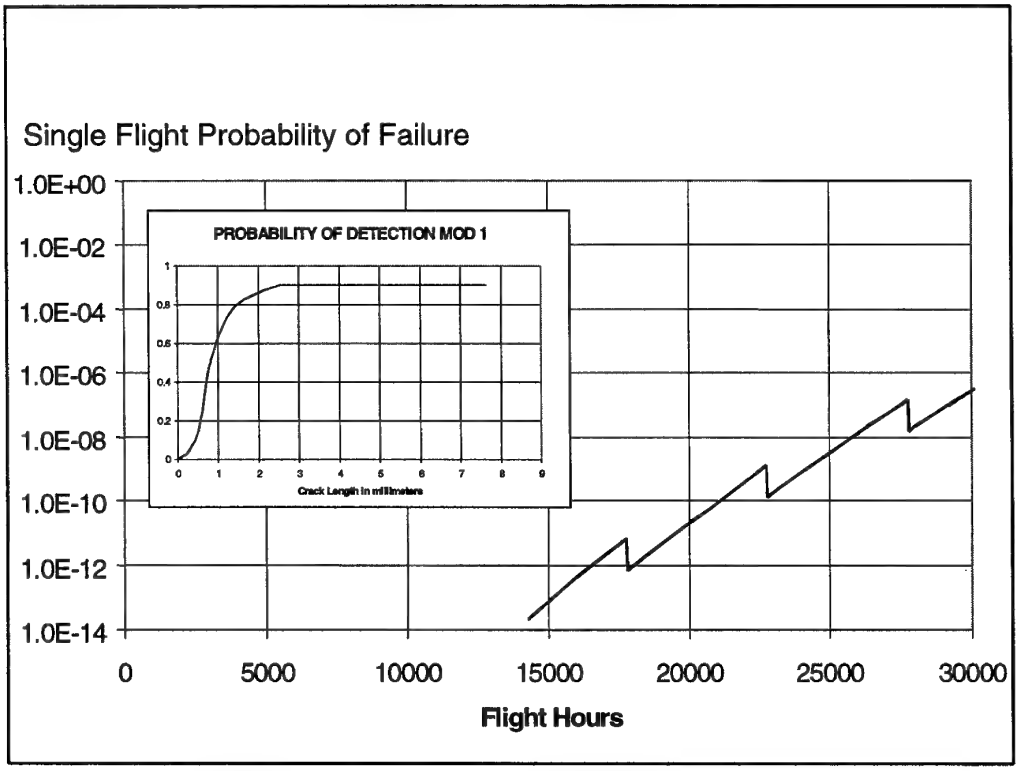


Figure 14 Risk with POD Modification 1

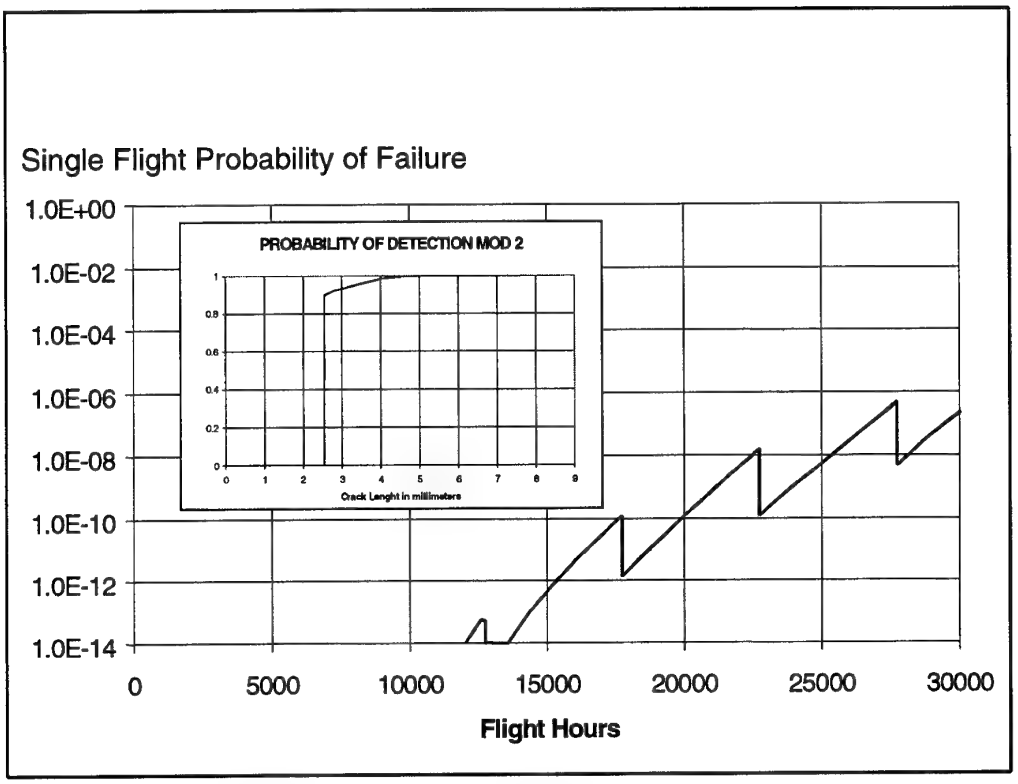


Figure 15 Risk with POD Modification 2

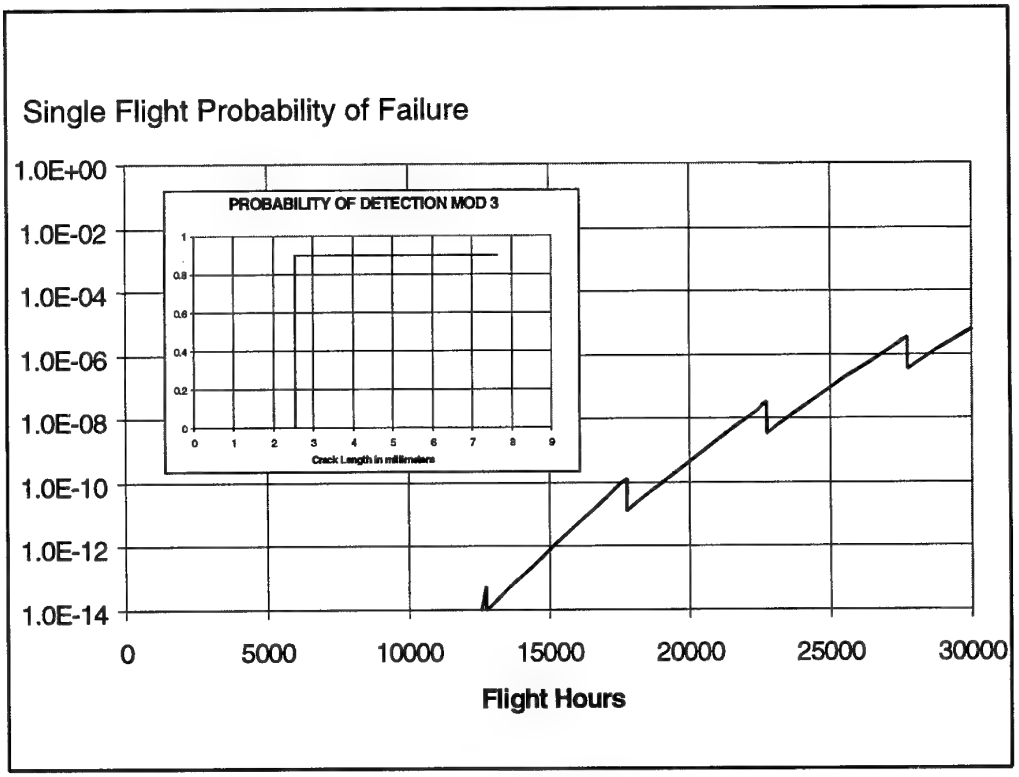


Figure 16 Risk with POD Modification 3

Factors Influencing Eddy Current PoD in the Field Environment

Don Hagemaier
Boeing, Phantom Works
2401 E. Wardlow Rd.
Mail Code C071-0013
Long Beach, CA 90807-5309, USA

SUMMARY

The probability of detection (PoD) is defined as the probability that, using a specific inspection procedure, a trained inspector will detect a flaw of a certain specified size (a_{det}).

Presented are those factors which influence the eddy current PoD in the field environment, i.e., on-aircraft inspections. Generally, these factors tend to lower NDI performance below that expected on the basis of capability as demonstrated in a laboratory environment. Hence, strict attention must be placed on minimizing the influence the following factors have on the inspection reliability:

1. Human factors and qualified personnel.
2. Access to the inspection area.
3. Inspector working to a specific validated written procedure.
4. Equipment variability.
5. Measurement repeatability.
6. Detectable crack size.
7. Signal-to-noise ratio.
8. Reference standards.

All of these factors must be considered and accounted for to have a reliable written inspection. The most frequent cause for unreliable NDI performance, as observed by Rummel [1], is that of improper NDE engineering. There is a "process", to arrive at reliable inspections. This process consists of the following steps:

1. Perform damage tolerance analysis of the area.
2. Marked-up engineering drawing showing crack location/orientation and crack growth curves.
3. NDT engineers determines the materials involved and the thickness of the structure.
4. Potential NDT methods are selected based on access and a_{det}
5. Simulated structure is designed and fabricated.
6. EDM notches of various sizes are fabricated in the reference standard.
7. Determine preliminary procedure and a_{det}
8. Finalize procedure and verify on operational aircraft.
9. Procedure reviewed by operator manufacture Working Group and Regulator prior to release.
10. Release and revise as necessary.

The most important point is determining the minimum detectable crack size and establishing the inspection threshold "A" which provides two or more inspections before the crack grows to a_{inst}. The inspection threshold "A" shall provide a signal-to-noise ratio of 3 to 1 or better.

INTRODUCTION

In-service aircraft checks are devised in order to detect degradation which might lead to premature failure. Experience gained in recent years with modern, pressurized airliners has emphasized the importance of maintaining a high level of structural integrity, particularly through vigilance against fatigue-type deterioration and stress corrosion. This, in turn, has emphasized the importance of inspection programs under which aircraft may attain long, safe service lives. Such programs must cover systematically the primary structure and structural joints, and must give attention to hidden areas and members subject to repetitive cyclic loads.

Paper presented at the RTO AVT Workshop on "Airframe Inspection Reliability under Field/Depot Conditions", held in Brussels, Belgium, 13-14 May 1998, and published in RTO MP-10.

To ensure the structural integrity of the older transports in service, the manufacturers developed an Supplemental Inspection Document (SID)Program for aging aircraft. A similar damage tolerance program, the Airworthiness Limitations Instructions (ALI), has been developed for new aircraft

AIRCRAFT INSPECTION PROGRAMS

Supplemental Inspection Program (SID) [2]

The structural integrity of aging aircraft is ensured through the FAA mandated Supplemental Inspection Document (SID) program. The SID's identify the principal structural elements (PSE's) on each aircraft. SID's also identify the inspection methods and procedures associated with each PSE. Briefly, this is an inspection program to supplement or adjust existing structural programs, as required, to ensure the continued safety of older aircraft.

Airworthiness Limitations Instructions (ALI) Document [3]

FAA Advisory Circular No. 121-22 (January 12, 1977) was developed to facilitate communication between FAA, operator, and manufacturer and provide the necessary guidelines for establishing and conducting a Maintenance Review Board (MRB) on newly manufactured aircraft, powerplant or appliance to be used in air carrier service, in order to develop the initial maintenance and inspection requirements for transport category aircraft.

Each new model aircraft will have its own Airworthiness Limitations Instructions (ALI) Document. The ALI document specifically addresses those items which have been identified through the certification process as being either safe-life (life limited) or damage tolerant and meet the definition of being a Principal Structural Element (PSE).

FACTORS INFLUENCING EDDY CURRENT PoD

Because eddy current inspection has become the primary crack detection method, the author will discuss those factors influencing eddy current PoD in the field environment, i.e., on-aircraft inspections

Human Factors and Personnel Qualification

For the purpose of this discussion, optimum or ideal performance is the capability of a proven NDI procedure to detect a crack of a specified size when the procedure is carried out by a qualified technician. An NDI technician is said to be skilled when he is qualified to carry out an inspection involving knowledge, judgment, and manual deftness, usually acquired as a result of long training, whereas an unskilled technician is not expected to do anything that cannot be learned in a relatively short period of time. The need for qualified NDT inspectors is well recognized throughout the NDT community. It is especially important for the SID/ALI programs because the person must be familiar with aircraft structure, must be trained in the applicable method, and must be proficient in following detailed written procedures.

FAA document FAR 121.371 (Required Inspection Personnel) clearly states: "(a) No person may use any person to perform required inspections unless the person performing the inspection is appropriately certified, properly trained, qualified, and authorized to do so; and (b) No person may allow any person to perform a required inspection unless, at the time, the person performing that inspection is under the supervision and control of an inspection unit."

Concerning technician performance, Rummel [1], states: " Errors in performance by skilled operators may be classified as: *Systematic* Error (consistent offset from ideal performance); *Errors In Precision* (consistent, but random, variation in performance about a norm); *Sporadic* Errors (an occasional occurrence varying significantly from the norm). *Sporadic* errors are usually associated with lack of motivation, boredom, fatigue, and monotony. *Errors in precision* can be caused by slight variation in processing, by inexperience of operator, or by a shift in decision criteria usually due to a lack of confidence. *Systematic* errors may be due to

a difference in skill and /or decision criteria input by the operators; or may be due to differences in equipment or calibration standards.

According to Gordon Dupont [4] there are the "dirty dozen" or 12 most common causes of a maintenance person making an error in judgment which results in a maintenance error. These same 12 can be applied to the NDI inspector. They are::
1. Lack of communication (never mind what the procedure says),
2. Complacency0 (constant repetition can cause error in judgment),
3. Lack of knowledge (poor training or outdated material)),
4. Distraction (losing track of where you're at),
5. Lack of teamwork ("I can do it myself"),
6. Fatigue (60-hour week),
7. Lack of resources (one instrument for three inspectors),
8. Pressure ("hurry up or we're going to be late"),
9. Lack of assertiveness ("we will correct it some day"),
10 Stress ("you want it when ? "),
11. Lack of awareness (" I don't care about the consequences, get it done"),
12. Norms (peer pressure).

Access to Inspection Area

Most fabrication NDI/PoD studies are conducted in a laboratory setting which generally matches the production environment in which it will be conducted. On-aircraft inspections can require certain steps be conducted prior to inspections. Some of these requirements are; remove paint, open access doors, remove auxiliary components (seats, insulation, lavatories, ducting, carpeting, etc.) to gain access to the area or part to be inspected. These necessary requirements add time delay and costs to the operators but they are necessary for a reliable inspection. Spencer [4] at Sandia National Laboratories, conducted a round robin study of an eddy current experiment for first layer crack detection. To simulate a realistic experiment, half of his specimens were painted and half were bare aluminum. He reported that the effect of inspecting through paint (0.003 to 0.005 in.) thick is often a decrease in the PoD. However, this effect is due to the difficulty in properly centering the probe over the rivets. Techniques that give the operator signal feedback that can be used to assure proper centering are effective in eliminating paint as a reliability factor.

When performing inspections on the crown of the aircraft, safety harnesses or platforms are required so that the inspector does not fall to the ground. The inspector cannot perform a reliable inspection if he is continually concerned about falling off the structure. The inspector can easily slip when the structure he is standing on is wet with oil or water vapor.

There are times when the inspector must enter the wing tanks. The tank must be drained and purged prior to entry plus a air vent tube or hose must be supplied to avoid CO_2 poisoning. In some cases, the man in the tank will manipulate the probe while another man outside the tank watches the instrument screen for crack indications. Additionally, the instrument must be precalibrated for liftoff due to internal paint thickness.

Validated Inspection Procedures

The most frequent cause for unreliable NDI performance, as observed by Rummel [1] is that of improper NDE engineering. In many cases, the NDI method selected is incorrect or was not qualified and controlled to the level necessary for the required discrimination. At Douglas Products Division (DPD) of Boeing, all NDI procedures are prepared by experienced NDE engineers, developed in the laboratory, and verified on operational aircraft. Hence, the lack of "up front" engineering is eliminated. The process is as follows:

Damage-tolerance analysis is performed for each PSE, and a marked-up engineering drawing and crack growth curves (Figure 1) of the component are submitted to the nondestructive testing (NDT) engineer. The location and orientation of any anticipated cracks are indicated on the drawing. The NDT engineer determines the materials involved and the thickness of the various parts making up the PSE. Potential nondestructive inspection (NDI) methods and

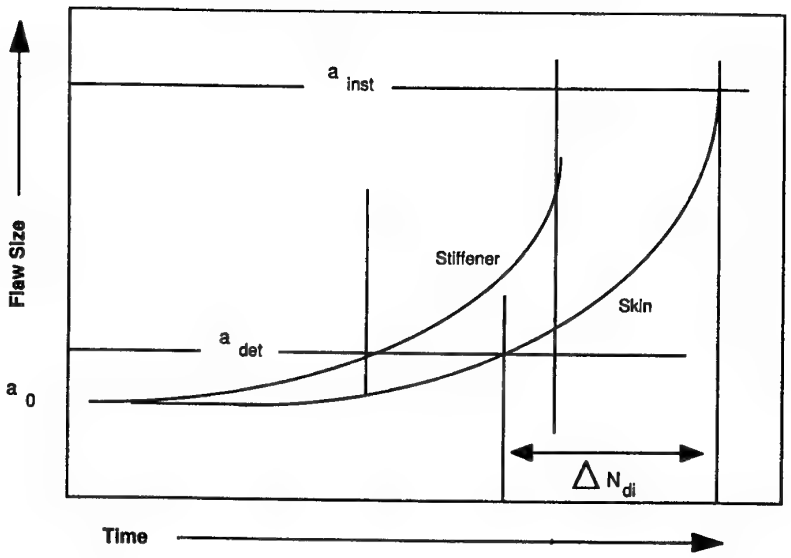


Figure 1. Typical Crack Growth Curves

techniques are then selected for inspecting the PSE [5]. For eddy current inspection, simulated structure is fabricated with electrical discharge machining (EDM) notches of different sizes. These notched specimens are then used to work out preliminary procedures and different detectable flaw size for each PSE. Obviously, the detectable flaw size must be less than the instability flaw size for each PSE. The procedures are finalized and then verified on operational aircraft. The verification procedure provides a means of detecting constraints that are not obvious from drawings or sketches.

It also defines access and/or removals required to perform the desired inspection. Finally, the inspection procedures are reviewed by the operator manufacturer working groups (for each model aircraft), and regulator prior to release.

Included in the procedure is a descriptive paragraph and illustration, along with specific details explaining the exact location of the PSE area within the major assembly. Illustrations are provided in the procedure which show not only probe placement onto the reference standard, but specific screen presentations which should be achieved during calibration, as shown in Figure 2. The procedure follows in a step-by-step manner complete with illustrations of the structure to be tested along with the orientation and scanning direction of the probes on the specific part to be examined, with illustrations of the flaws that are to be found.

Equipment Variability

Most SID/ALI inspections require the use of eddy current equipment. These inspections will be conducted at maintenance bases located throughout the world. Therefore, a variety of equipment will be used. In each inspection procedure, the specific equipment used is identified (see Figure 2). However because each operator may not have this specific equipment, an "or equivalent" statement is used.

The problem is to determine equivalency between two similar instruments from different manufacturers or two identical instruments from one manufacturer. This is especially true for similar eddy current probes Figure 3 shows photoinductive field maps of "identical" 2 MHz absolute probes, as evaluated by Moulder [6]. It is very obvious that the output from similar probes is "not identical".

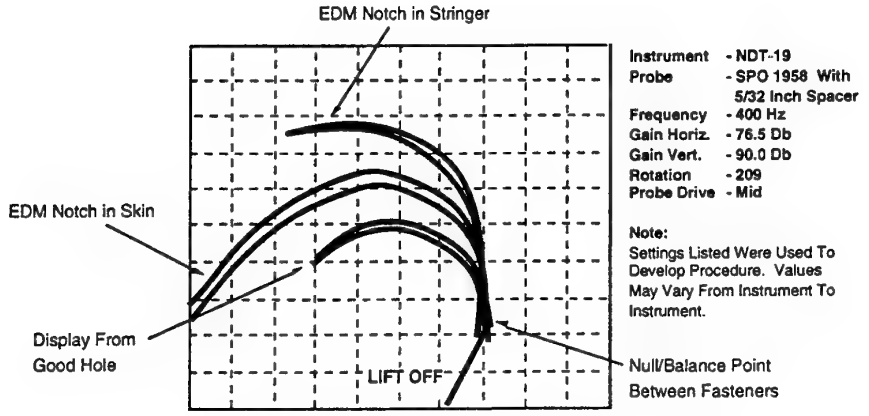


Figure 2. Typical Calibration Figure

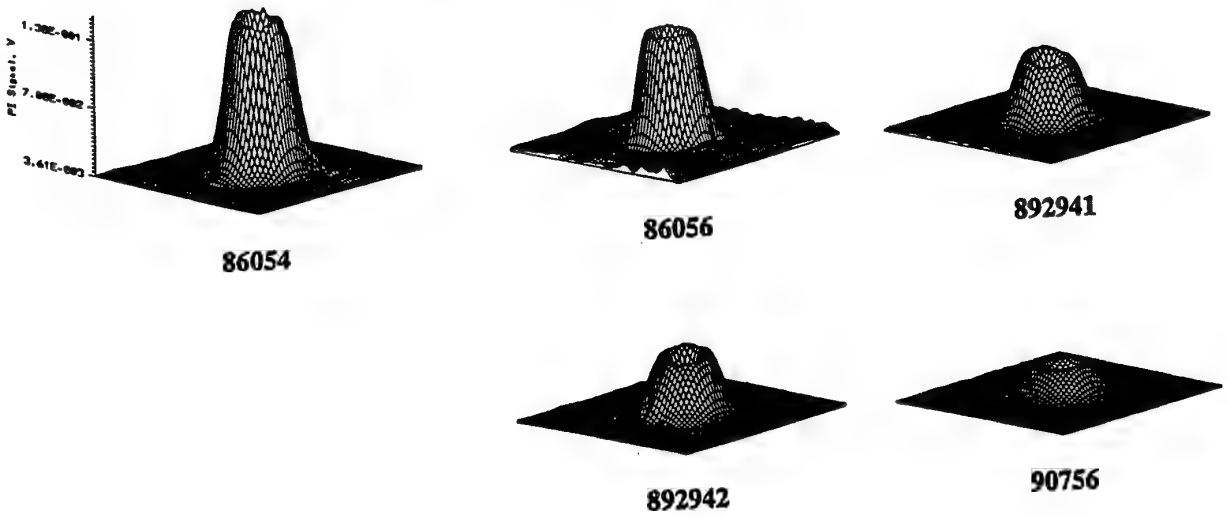


Figure 3 Photoinductive Field Maps of 2 MHz Absolute Probes (after Moulder)

Probe 86054 achieves a 90% PoD at 0.75 mm whereas probe 90756 almost achieves 90% at 1.5 mm.

Unfortunately, few manufacturers have a quantitative calibration procedure that can be used to determine equivalency or repeatable performance. Hence, qualitative methods must be used. These qualitative procedures generally entail using primary or secondary reference standards to calibrate the eddy current instrument prior to and periodically during inspection of a particular part. Results from similar or identical equipment may be compared by use of simulated-defect (electrical discharge machined, EDM

notched) reference standards. Usually, the reference EDM notch size is representative of the detectable crack size.

Measurement Repeatability - Despite all efforts to ensure repeatability, experimental measurements of eddy current flaw-signal amplitudes are never exactly the same, in the strict mathematical sense, over a set of repeated scans of the same flaw. Instead, the signal amplitudes thus obtained form a distribution of values ranging from a minimum to a maximum and having some mean average value. If one were to calculate the number of times a given amplitude was observed divided by the total number of scans and then plot the resulting data as a function of signal amplitude, the curve obtained would be the probability density function (PDF) for the signal amplitudes from that particular flaw size. A similar PDF for noise or background signals can be defined in much the same way.

Two such PDFs, one for the eddy current flaw signal and the other for the noise, are shown in Figure 4 [7,8].

In an inspection situation, one would hope that the PDF for flaw signals would lie well to the right of the PDF for noise so that a given signal amplitude could be unambiguously interpreted as either a flaw signal or noise. In such an ideal case, most flaws would be detected, and there would be no false alarms from background signals that appear to indicate the presence of a flaw. In practice, this ideal situation is realized only for very large flaws in the presence of very weak noise signals. However, when testing for small flaws, the flaw signals and noise overlap to some extent, as indicated in Figure 5. It is the extent of the overlap or, more precisely, the area under the PDF curves in the overlap region, that determines the reliability of the inspection. Note that the 0.635 mm (0.025 in.) crack/notch signals are buried in the noise at 5 dB and hence undetectable [9]. The signal-to-noise ratio for the 32 test sights are listed in Table 1. The 1.27 mm (0.050 in.) crack/notch signals have a signal-to-noise (SNR) ratio of about 3 to 1. The noise amplitude from the uncracked fastener locations does not exceed 5 dB. Hence, a flaw gate could be set at 15 dB for a reliable inspection.

Fortunately, the newer flying-dot eddy current impedance- plane instruments give a clear separation between noise (lift-off and no-crack) and flaw signal, as shown in Figure 2

Where signal and noise PDFs overlap (Figure 4) to a significant degree, both Type I and Type II errors will occur. Hence, a decision must be made to establish a threshold value (a_{det}) favoring either Type I or Type II errors.

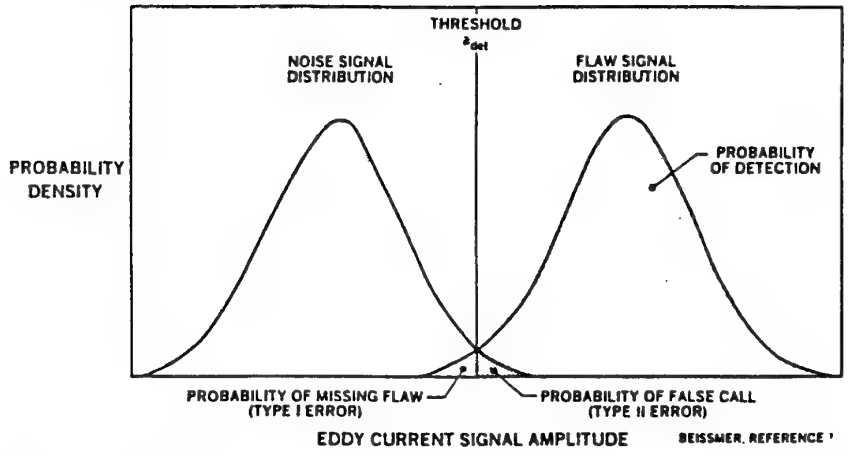


Figure 4. Probability Density Functions for Signals and Background Noise - Typical Case (After Beissmer)

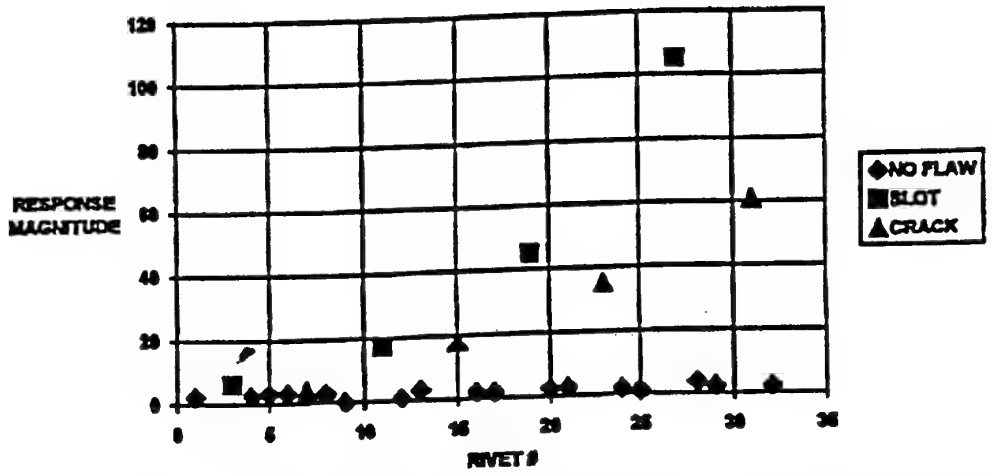


Figure 5. Signal Response Magnitude for Surface Cracks/Notches Under Aluminum Rivets (After Sheppard)

Table 1. SNR for First Layer Cracks/Notches Under Aluminum Rivets

Flaw Size	Slot or Crcak	SNR (dB0
0.635	Slot	5.0
0.635	Crack	1.5
1.27	Slot	14.0
1.27	Crack	14.7
1.90	Slot	22.6
1.90	Crack	20.5
2.54	Slot	30.0
2.54	Crack	25.0

The area under the noise PDF (to the right of a_{det} value) is then the probability that Type II errors will occur. At the same time, the choice of a particular a_{det} value will also determine the probability of flaw detection, because the area under the flaw signal PDF (to the right of the a_{det} value) is the PoD. The a_{det} threshold value also determines the occurrence of Type I errors that are equal to the area under the signal (flaw) PDF (to the left of the threshold value). Thus, the extent of overlap of the flaw signal and noise PDFs and the choice of a threshold amplitude for flaw detection play critical roles in determining the reliability of each inspection method or technique.

Detectable Crack Size

In practical applications, an NDI limit for detectable flaw size, a_{det} , is usually specified; this is a crack length (a) corresponding to a high probability of detection. Detectable crack size is different for each inspection method and PSE. Although the detectable crack size is different for each method, the probability of detection (PoD) in the SID/ALI programs are considered to be 0.9 regardless of the method chosen [2]. However, the method chosen will govern a_{det} and hence establish inspection start points and intervals.

A primary NDI method and at least one alternate method are developed for most PSEs. The primary method is the most sensitive method; i.e., it can detect the

smallest crack and it gives the largest crack growth interval, ΔN_{di}.

The primary interest in the aircraft structural inspections is the probability of positive detection. Because the PoD curve graphically depicts discrimination capability, it is a convenient way to compare inspection process performance. The PoD curve does not, however, provide an indication of the calibration performed to establish the baseline, the acceptance criteria imposed on the process, or the level of incorrect rejections (false calls) inherent in the application. The common denominators for both NDI performance and modeling the performance of a specific technique are (1) the signal and noise response distribution generated by application of the technique, and (2) the acceptance criteria applied to the decision process.

A PoD curve typically reflects all the variations in signal-to-noise response and discrimination level shown in Figure 6. A continuing variation in signal-to-noise response is reflected by variation in the discrimination level (threshold) along the PoD.

Where the NDI response (signal) distribution from a flaw is coincident with the process noise signals, there is no discrimination and the inspection is not valid. This is true for the 0.635 mm (0.025 in.) crack & notch in Figure 5 [9].

In order to achieve successful detection, NDT engineers determine a minimum detectable flaw size for each inspection. This size is obtained from the laboratory demonstration and is defined as a_{det} in Figure 6. At this minimum threshold, there may be some Type I and Type II errors. Also, this a_{det} threshold is developed in the laboratory where conditions are optimum. To be assured of positive detection in the field, the engineer chooses a slightly larger crack size threshold, i.e., threshold "A" in Figure 6. At this threshold (SNR=3:1), there is good separation between flaw signals and noise, resulting in a reliable inspection. In addition, decision criteria (crack versus no crack) are clearly defined. Reference Figure 6.

Before eddy current or ultrasonic inspection is performed, the instrument is calibrated using an EDM notch of the appropriate size. This notch size is equal to the "A" determined in the laboratory. The "A" value must provide a signal-to-noise ratio (SNR) of 3:1 or better and be less than a_{inst} at limit load. The "A" value, in Figure 5, is 1.27 mm (0.050 in.) notch/crack length.

Should the inspector obtain a positive flaw response, the following criteria are used to make a determination of inspection results:
1. For cathode-ray tube eddy current, any indication that exhibits the same relative phase angle and an amplitude equal to or greater than the reference notch is considered a crack.
2. For meter eddy current, any indication that exhibits an amplitude equal to or greater than the reference notch is considered a crack.

90% Reliability at a 95% Confidence Factor [1,2,7]
With the advent of the damage tolerance approach and maintenance philosophy, there was a decided need for achieving greater effectiveness and reliability in the application of NDI. Through the developing years of NDT, there was no parallel development of the ability to express NDI results in discrete quantitative terms. Previously, the question was; "How small a crack can be detected?" Now the question is; "How large a crack can be missed?"

The NDI goal is to establish a value for a_{det} and have an inspection system that produces low error modes: Type I (failure to find a crack when one is present or is smaller than a $_{det}$) and Type II (indicating a crack when none is present). The two positive modes are: (indicating a crack when one is present) and (indicating no crack when none is present).

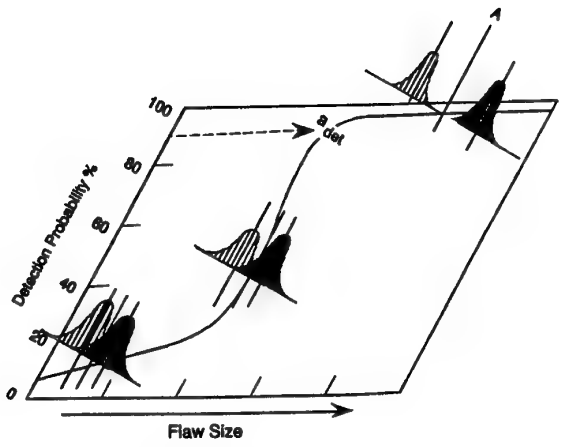


Figure 6. Interaction of Signal/Noise Discrimination and the PoD

The probability of detection (PoD) is defined as the probability that, using a given inspection procedure, a trained inspector will detect a flaw of a certain specified size if it exists. The flawed specimens are mixed with a number of unflawed specimens and subjected to inspection using production equipment and personnel.

Confidence level means that the more we know about anything the better our chances are of being right. For a large sample size, 100 percent confidence can be gained when a measurement coincides with the true value. For a small sample size, confidence level is established in terms of actual sample size and the success or failure rate within that sample. A confidence level is then based on history repeating itself and therefore specifies the percentage of the time we expect to be correct. No information is conveyed regarding the total number of flaws that will be found in a demonstration program.

To demonstrate that a 0.5 mm (.125 in.) long crack can be found at 90/95, the inspector must find that crack 29 times out of 29 attempts. He is not allowed to miss that crack even once in 29 times. If it was missed once, then on the second try, he must find it 45 out of 46 attempts., or with two misses and 59 successes in 61 trials, and so forth. Some people think that a 50% confidence limit means a 50/50 chance of success. In order to demonstrate 90% reliability at 50% confidence level, that 0.5 mm (.125 in.) long crack must be found 7 times out of 7 attempts. If it is missed once in 7 tries then it must be found 16 times out of 17 attempts, and so on. Even a 50% confidence bound is a relatively high level of confidence statistically speaking. For comparison, see Table 2.

To achieve the 90% PoD with 95% confidence, there must be X successes in N False-Call rate. See Table 3 trials. The design flaw size must be the largest in these N trials without exceeding the false call rate (Table 3).

Table 2
Number of Successes Required (Reliability-Confidence)

For 90% - 95%	For 90% - 50%
29 in 29 Trials	7 in 7 Trials
45 in 46 Trials	16 in 17 Trials
59 in 61 Trials	25 in 27 Trials
72 in 75 Trials	34 in 37 Trials
etc.	etc.

Table 3.
Demonstration Requirements

Successes	Trials	False Calls
29	29	3
45	46	5
59	61	7

For an FAA funded program, fatigue cracks were generated in panels simulating a fuselage lap joint. The program was initiated to determine the minimum detectable crack size for surface eddy current inspection around flush-head rivets. The panels were taken to a variety of airline maintenance bases for test and evaluation. The study was managed by the FAA Aging Aircraft NDE Validation Center at Sandia National Laboratories in Albuquerque. The results [10],indicated a 90/95 PoD of 2.54 mm (0.100 in.) from the shank of the rivet A similar study was later conducted at the Validation Center by various eddy current equipment manufacturers [4]. The results of this study were more encouraging in that at least one of the instruments was capable of detecting cracks 1.0 mm (0.040 in.) in length.

Signal-to-Noise Ratio

Where the NDI response (signal) distribution from a flaw is normal and where process noise is well separated from the signal, the inspection has high specificity for discrimination of the signals (good signal-to-noise ratio). See the upper portion of Figure 6. Where the NDI response (signal) distribution from a flaw is coincident with the process noise signals, there is no discrimination and the inspection is not valid (lower portion of Figure 6). It is obvious from Figure 7 that if the threshold is set to detect very small cracks then the Type I and Type II errors overlap (small cracks are missed and false calls are made). However, Figure 6 shows that if the threshold is set at "A" then there is good discrimination between noise and signal.

Figure 5 illustrates the signal to noise ratio for a surface eddy current test [9].The specimen contained 32 fastener locations. Four locations contained EDM notches and four contained fatigue cracks. The notches and cracks ranged from 0.635 mm (0.025 in.) to 2.54 mm (0.100 in.) in length in the first layer aluminum sheet at the fasteners. The voltage from the unflawed locations was 5 mV maximum. The signal amplitude from the 0.635 mm (0.025") notch and crack were juxtaposed with the noise and hence, no discrimination. However, for the 1.27 mm (0.050") notch and crack, the signal voltage was 15 mV. This gives a signal-to-noise ratio of 3 to 1 which is adequate for inspection with this technique.

Reference Standards - Electrical discharge machined (EDM) notches of the appropriate size are placed in simulated structural reference standards for the purpose of: 1) determining the detectable flaw size for each inspection; 2) setting up the eddy current equipment prior to inspection and at periods during the inspection, and 3) evaluating the relative size of flaw signals in relation to the acceptance/rejection criteria. Generally, the size of the notch yields a signal-to-noise ratio of 3 to 1 for a particular equipment/probe combination used for a specific inspection.

"Calibration" of the eddy current system is frequently accomplished by adjusting the eddy current instrument to produce a predetermined response to a known size notch [9].

Automated Eddy Current Scanning For the past few years, the author and his coworkers have been developing automated eddy current scanning for corrosion [11] and crack detection [12]. It is felt that the plan-view scans add to the reliability of the inspections because the size, shape, and depth of corrosion is clearly shown. Also, the length and orientation of cracks are revealed in a permanent record. In addition, the automated scanning reduces the time required to perform inspections of complicated structure.

Typical results obtained for a 0.035 inch long fatigue crack under a flush-head aluminum rivet is shown in Figure 7. The first layer material was 0.040 inch thick 2024-T4 aluminum. These results were obtained using the NASA LaRC, Self-Nulling Rotating Probe System. This system has achieved a 90/95 PoD for a 0.032 inch long first layer fatigue crack.

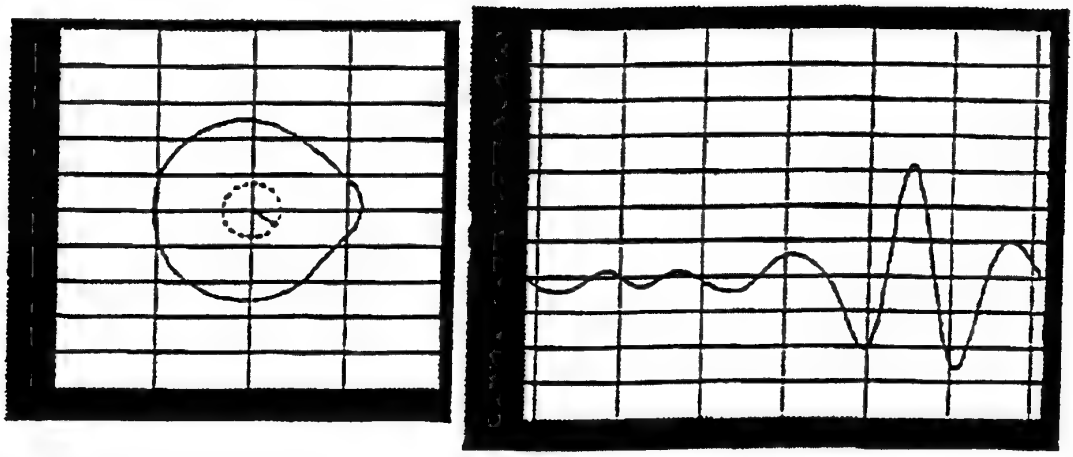


Figure 7. Real-Time Display of Rotating Probe System - 0.035 inch Crack Under Al Rivet

REFERENCES

1. Rummel, W., "Human Factors Considerations in the Assessment of Nondestructive Evaluation (NDE) Reliability," Review of Progress in Quantitative Nondestructive Evaluation, Vol. 3A, ed., D. O. Thompson and D. E. Chimenti, 1984. Plenum Press, New York, N Y, and London, UK.

2. Abelkis, P., Hagemaier,D., and Harmon, M., "Supplemental Inspection of Aging Aircraft," Materials Evaluation, July, 1986

3. Hagemaier, D., and Wilson, D., "Organizing NDT in the Aerospace Industry: The Airworthiness Limitations Instructions (ALI) Document", Materials Evaluation, April, 1994

4. Dupont, G., "The Dirty Dozen Errors in Maintenance," FAA, Human Factors Issues in Aviation Maintenance and Inspection, March 12-13,1997, San Diego, California.

5. Spencer, F., "Detection Reliability For Small Cracks Beneath Rivet Heads Using Eddy Current Nondestructive Inspection Techniques," FAA Aging Aircraft NDI Validation Center, Sandia National Laboratories, Albuquerque, NM, 1997.

6. Hagemaier, D., and Hoggard, A., "NDT Technology as It Relates to Aging Aircraft," Materials Evaluation, Dec. 1993

7. Moulder, J., "Eddy Current Probes," FAA Center for Aviation Systems Reliability, Iowa State University, I996.

8. Beissmer, R., et al., "Exploratory Development of Advanced Surface Flaw Detection Methods," AFWAL-TR-4121, 1984. Materials Laboratory, Wright Patterson AFB, OH.

9. Government/Industry Workshop on the Reliability of Nondestructive Inspection - Proceedings, SA-ALC/MME 76-6-38-2, 1978.

10. Hagemaier, D., and Kark, G., "Eddy Current Detection of Short Cracks Under Installed Fasteners," Materials Evaluation, January 1997.

11. Spencer,F., and Schurman,D., "Reliability Assessment at Airline Inspection Facilities, Vol. III: Results of an Eddy Current Inspection Reliability Experiment," DOT/FAA/CT-92/12,III,May 1995.

12. Hagemaier, D., Jappe, W., Wilson, D., and Wood, N., "Improved NDI Techniques for Aircraft Inspection," The First DoD/FAA/NASA Conference on Aging Aircraft, Ogden, Utah, July 1997.

NDT reliability estimation from small samples and in-service experience.

Dr David A Bruce
Structural Materials Centre
DERA Farnborough
Ively Road, Farnborough, Hampshire GU14 0LX
United Kingdom

Optimisation of an inspection strategy to provide acceptable safety at minimal cost requires a knowledge of the reliability of the inspection procedures which could be used. A methodology for assessing inspection reliability, characterising the inspection process by a 95% confidence level probability of detection (POD) curve estimated from artificial trials, has become the standard approach. This method works satisfactorily for straightforward inspection situations where the POD curve can be estimated from a large database, but application of similar methods to airframe inspection suffer from the prohibitive cost of obtaining the reliability curve from realistic trials. Where there is limited data available to determine the reliability, the inbuilt conservatism of the standard method leads to wholly unrealistic estimates for the POD curve which in turn give rise to unacceptably short inspection intervals and excessive maintenance costs. It may be possible to deduce inspection reliability from in service inspection data, although the diversity of inspection situations suggests that there will still be a very limited amount of information available from which to estimate the reliability for a particular inspection task. In this paper the effect of the inbuilt conservatism inherent in the standard method of POD assessment will be demonstrated. Alternative approaches to the prediction of NDT performance will be compared to establish the minimum requirements for inspection data in order to achieve specified safety levels. The possibility of using techniques based on Bayesian inference to provide an optimal prediction of reliability which can be refined as further information is acquired will be described. The effects will be demonstrated using simulated data based on real inspection reliability trials.

Paper presented at the RTO AVT Workshop on "Airframe Inspection Reliability under Field/Depot Conditions", held in Brussels, Belgium, 13-14 May 1998, and published in RTO MP-10.

1. Introduction

Non-Destructive Testing (NDT) is used on virtually every type of military or commercial aircraft whatever the original design philosophy. Whenever NDT is used on aircraft primary structure to detect potentially critical defects, the reliability of the inspection method used becomes one of the principal factors determining the safety level at which the aircraft operates. A common rule of thumb assumes that a defect should be inspected at least three times during the period in which it grows to the maximum acceptable size. This is based on the assumption that the probability of detecting the defect is 90% for each of the three inspections. This leads to a safety level, the probability of missing the defect completely, of 1 in 1000. This level is currently accepted by airworthiness authorities.

Optimisation of an inspection strategy to provide acceptable safety at minimal cost requires a knowledge of the reliability of the inspection procedures which could be used. The parameter which is usually used to describe inspection reliability is the "Probability of Detection" or POD. The actual figure quoted is not an estimate of the true probability, but is a lower bound calculated at a desired confidence level[1]. For detection of growing fatigue cracks or similar defects it is useful to know the reliability of inspection methods as a function of the defect size. A methodology for assessing inspection reliability, characterising the inspection process by a 95% confidence level POD curve estimated from artificial trials, has become the standard approach[2].

This method works satisfactorily for straightforward inspection situations where the POD curve can be estimated from a large database, such as occurs in engine disk inspection for example. Application of similar methods to airframe inspection suffer from the prohibitive cost of obtaining the reliability curve from realistic trials. Where there is only a limited amount of data available to determine the reliability of an inspection method, the inbuilt conservatism of the standard method leads to wholly unrealistic estimates for the POD curve which in turn give rise to unacceptably short inspection intervals and excessive maintenance costs.

In order to overcome the problem of providing realistic data, it may be possible to deduce inspection reliability from in service inspection data. Although many inspections are carried out and many defects found, the diversity of inspection situations including access, geometry and equipment variations suggest that there will still be a very limited amount of information available from which to estimate the reliability for a particular inspection task. A more efficient procedure for interpreting the limited data available and predicting the reliability of field inspections is therefore essential if the cost savings arising from greater dependence on NDT are to be realised.

In section 2 of this paper the effect of the inbuilt conservatism inherent in the standard method of POD assessment will be demonstrated. The possibility of using real inspection data from in-service inspections will be explored in section 3. Finally alternative approaches to the prediction of NDT performance will be compared to establish the minimum requirements for inspection data in order to achieve specified safety levels. The possibility of using techniques based on Bayesian inference to provide an optimal prediction of reliability which can be refined as further information is acquired or as a means of tailoring generic inspection reliability estimates to a particular inspection situation will be described. The effects will be demonstrated using simulated data based on real inspection reliability trials.

2. POD estimates from small samples

2.1 Inherent conservatism of POD procedures

2.1.1 Probabilities of Detection p_t and p_m.

The underlying assumption in a probabilistic analysis of any process, in this case a particular inspection, is that if a large number of independent trials were carried out, the proportion of these trials yielding a particular outcome would be a well-defined fraction called the true probability, p_t.

True probabilities represent the expected outcome of an infinite number of trials, hence it is impossible to measure p_t exactly. The best estimate of the true probability is given by the fraction of trials which yielded the outcome in question during a real, finite series of independent trials. Throughout this paper we will use n to designate the number of trials carried out on specimens containing defects. Colloquially, inspections which successfully detect defects are referred to as "hits" and those which are unsuccessful as "misses". As these terms form a useful shorthand we will use them freely throughout this paper. If the total number of trials is denoted by n and the number in which the defect was detected is denoted by h (for hits), then the best estimate for p_t is simply the mean probability obtained in the trials, $p_m = h / n$.

The importance of measuring probabilities of detection is, of course, to be able to predict the performance of an inspection strategy. From the mean probability p_m, we can obtain the best, unbiased, estimate for the number of defects which would be detected in a future set of n_2 trials. This is simply given by the product $p_m \times n_2$. Ultimately, for airworthiness purposes, the quantity of interest is the chance of missing a defect completely during its growth to critical size. The best estimate of this is given by $(1 - p_m)^{n}_{2}$ where n_2 represents the number of inspections during the defect growth period, typically three.

The limitation on using p_m to predict the outcome of future tests is that a simple knowledge of p_m does not contain any information about how accurate our estimate of the true probability is, and hence how far out the estimate of the number of hits and misses might be. For this reason, the mean probability p_m is not used as a measure of NDE reliability when limits on this are specified.

2.1.2 Confidence limits and the "POD" p_α.

The procedure which has been adopted in NDE reliability studies, following the recommendations of guidelines published by the American Society for NDT (ASNT)[3,4], is to specify the "Probability Of Detection at a specified Confidence Level." The confidence level is usually taken to be 95%. We denote this probability by p_α, where α indicates the confidence level. The "POD" p_α is not, in fact, a probability. It is a lower bound or confidence limit on the estimate of the true probability p_t. The confidence level α means that if p_t is actually lower than p_α, there is a probability of only $1 - \alpha$ that the data obtained in the experiment could have resulted from n independent trials.

There are several procedures which may be used to determine the value of p_α resulting from a particular experiment, all of which require a knowledge of the expected distribution of results. If a series of identical trials (inspections) is carried out, the expected outcome is governed by the binomial distribution. For a single trial, the probability of a hit is assumed to be p_t. The probability of obtaining exactly h hits in n trials, denoted

$p(h,n,p_t)$, is then given by the binomial probability function

$$P(h,n,p_t) = \binom{n}{h} p_t^h (1-p_t)^{(n-h)}$$

where $\binom{n}{h}$ is the usual binomial coefficient n!/h!(n-h)!.

Having obtained a value of h hits from a series of n trials we can now calculate a confidence level, α, for any estimate, p_α, of p_t. The confidence level which we ascribe to the probability p_α is the probability of obtaining at least h hits from n trials given that the probability of obtaining a hit in a single trial is p_α. Using the above probability function, α is then defined as

$$\alpha = 1 - \sum_{r=h}^{n} p(r,n,p_\alpha)$$

The value for the desired POD can be obtained from this equation by setting the required confidence level α, estimating p_α and then adjusting this estimate iteratively until the equation is satisfied. This conceptually simple procedure is completely general and can easily be carried out by computer. The POD p_α is necessarily a conservative estimate of the capability of the inspection technique. A knowledge of p_α alone does not allow the expected performance of an inspection technique to be predicted. The greater the confidence level required the greater the discrepancy between predictions based on p_α and the best, unbiased predictions calculated using p_m.

2.1.3 Variation of p_α with number of trials

The requirement to demonstrate a certain level of POD at a given confidence level places severe constraints on the experiment which must be carried out in order to verify the reliability of the technique. The crucial point is that the difference between p_t ($\approx p_m$) which represents the actual performance and the POD estimate p_α is strongly dependent on the number of trials, n, which were used to establish the POD. If n is large and there are a large number of hits and misses, then the difference $p_m - p_\alpha$ is proportional to the standard deviation of the number of misses expected for n trials s_n, since

$$p_m - p_\alpha \approx \frac{z_\alpha \sigma_n}{n} = z_\alpha \frac{\sqrt{n p_t (1-p_t)}}{n}$$

where z_α is a constant derived from the normal distribution which depends only on α. For the particular case of the 95% confidence level, z_α = 1.645. For smaller numbers of trials, in particular for experiments which yield a small number of misses, $p_\alpha - p_t$ must be calculated from the exact expression for α. The difference between the actual performance under test p_m and the estimated value p_α sets an upper limit to the POD which can be measured in an experiment of a given size.

2.2 Expectation values of POD

To illustrate the effect of sample size on the measured POD for a given p_t, the expectation value of p_α can be calculated by averaging the p_α values corresponding to all possible numbers of hits, weighted by the probability of this outcome, $p(n,h,p_t)$. In figure 1 this variation of expected p_α with n is shown for an assumed p_t of 0.95 and various values of α. It can be seen that although the true probability of detection of a defect in the trials is 95%, the POD estimate p_α is lower than this, the discrepancy being considerable for the higher confidence levels and smaller sample sizes. The most common requirement for p_α is to demonstrate 90% POD at the 95% confidence level. It can be seen from fig 1 that although p_t is considerably above the desired value, this is unlikely to be validated at the 95% confidence level unless more than 100 specimens are used.

The effect becomes more pronounced as the actual reliability of the technique approaches the required p_α. In figure 2 the POD expectation values are plotted for a technique with a p_t of 0.92. Here it can be seen that although the technique is capable of detecting over 90% of the defects, it is unlikely to be possible to verify this to even the 80% confidence level on a set of 200 trials.

In most uses of PODs, the minimum acceptable POD and confidence level are specified. These have then to be verified by an experiment. In order to estimate the sample size required for an experiment to demonstrate a given p_α the above procedure can be reversed. Rather than using the expression for α iteratively to calculate p_α for fixed values of h and n, it can be used to calculate the minimum necessary value of h for fixed p_α and n. The resulting p_m then gives the minimum necessary value of p_t for the inspection process at which the verification experiment would be expected to be successful. If p_t is exactly equal to this minimum value, the verification exercise will have approximately a 50% chance of success.

The sample size can be increased by increasing the number of specimens, the number of inspectors or both. The most common and straightforward method is to aggregate the data over a large number of specimens with nominally similar defects. It is reasonable to do this provided the defects can be made sufficiently similar that the POD is identical for all of the defects used and hence gives a value which will describe the probability of detecting any defect of this type and size. If necessary, this assumption can be tested in the same way as the data from different inspectors is tested for homogeneity in the next section[5].

2.2.1 Combining data from several inspectors

Aggregating the data over a number of inspectors is more liable to introduce problems. In principle, if the above assumptions are correct and there is a unique p_t for a given type and size of defect which is independent of the operator carrying out the inspection, then it is quite permissible to aggregate the data over a set of inspectors before calculating p_α. In practice, however, it is often found that there are significant differences in performance between individual inspectors and sets of equipment. These may arise from differences in skill and experience on the part of the inspectors or from differences in the calibration and set-up procedures for the equipment. Whenever data is aggregated, whether over specimens, inspectors or both, to calculate a single p_α, a check should be made to ensure that the data does not conflict with the binomial hypothesis which underlies the calculations.

The hypothesis that the data obtained in a set of several series of inspections can be described by a binomial distribution can be tested by any one of a number of standard statistical procedures. The principal constraints are the numbers of inspectors and specimens and the reliability of the inspection technique used.

If a sufficiently large number, N_I, of inspectors have taken part each inspecting N_D identical defects, a standard test such as the chi-squared (χ^2) test or the Kolmogorow-Smirnov test for distributions may be used to check for homogeneity. The former is more widespread.

The use of χ^2 (or Kolmogorow-Smirnov) tests may not be possible when assessing methods with a high p_t since, unless a very large number of specimens are used, there will be too few inspectors missing appreciable numbers of defects to allow the shape of the distribution to be tested. It is possible, however, in this situation to test whether parameters derived from the experimental results are consistent with the binomial

hypothesis[6]. The distribution of numbers of misses, m_i, obtained by each of the N_I inspectors can be described by its variance. The variance is a measure of the spread of the results. If the inspectors are not operating with a uniform probability of detection the variance should be larger than the expected value for a homogeneous group of inspectors. If the data for the numbers of misses obtained by each inspector is distributed according to a binomial distribution, then the variance should be equal to the mean. The variance ratio, var / m should therefore have a value of unity. The test statistic $N_I \times$ var /m can be shown to be distributed as χ^2 over N_I - 1 degrees of freedom. Testing that the value of var /m is not significantly greater than unity can therefore be used to determine whether the population of inspectors contains outliers who are performing less well than the average. As with the χ^2 test, if the test statistic is significant the binomial hypothesis must be rejected.

Rejection of the binomial hypothesis implies that there are significant differences between the performances of the individual inspectors. The data obtained by different inspectors cannot therefore be aggregated to calculate a single p_α as implied in the simple analysis. If the reasons for the poorer performance of specific individuals can be established (as being due to equipment malfunction or calibration difficulties for example) then they can be removed from the sample and the remaining homogeneous set of inspectors used to establish a POD.

If, however, the reasons for the poorer performance of specific individuals cannot be established, then it must be concluded that the simple binomial model which describes the inspection by a single p_t is inadequate[7]. In this case the results of the POD verification trials must be treated very carefully. An individual p_α value can be computed for each inspector. This will give the correct confidence limit for the result of a series of trials carried out by that inspector. If the results of all inspectors are aggregated to calculate an overall p_α for the set of inspectors, then this will only give the correct confidence limit for a subsequent set of trials if, for each trial, the inspector is selected randomly as part of the inspection process. The p_α obtained from the aggregated results would not be appropriate to describe the outcome of repeated inspections by a single inspector.

3. Reliability determination from in-service data

3.1 Realism in reliability studies

The costs and time required for major reliability verification exercises imposes two severe limitations on the results, particularly when used for airframe structural inspections. Inevitably the cost of producing large numbers of specimens becomes prohibitive if the specimens are too complex. If fatigue cracks have to be grown artificially, it is extremely difficult to produce these in a controlled manner in anything other than simple plate or dog-bone specimens. Although these simple elements may later be incorporated artificially in a structure to mimic some aspects of the real inspection, it will always be difficult to reproduce with confidence the human factors which are expected to affect operator performance and hence inspection reliability.

A further important limitation of artificial trials is the expectation of the operators taking part. To ensure realism in the inspections it is essential that the specimens containing defects should be accompanied by a number of specimens without defects to prevent over-reporting. Although the number of specimens which do not contain defects is not fixed and does not enter in the POD analysis, it is a normal rule of thumb[4] that it should be at least as large as the number containing defects. In practice this seldom approaches the situation

in the field, where it may be assumed that the vast majority of elements inspected will be free from defects. The result may be that in the artificial trials, operators expect to find defects in many of the specimens and so may be more conscientious in looking for defects and ready to report indications. In the field, defects will be encountered only occasionally, possibly leading operators to be less acute in detecting marginal indications leading to a decrease in reliability. A similar effect may result from the additional cost and disruption caused by the detection of a defect in a routine inspection in the field. Whereas there is no cost associated with reporting defects in an artificial trial, there is a considerable responsibility in reporting a defect in an aircraft which might result in that aircraft and possibly others in the fleet being grounded until repairs can be affected. The effects of this pressure are not known but again it may be surmised that the operators will be less likely to report marginal indications as defects.

A possible solution to both the cost of providing sufficient specimens for reliability trials and to the accusations of lack of realism which have been levelled at past reliability studies is to assess the reliability of inspection methods from real, in-service inspection results[8,9]. A limitation to this approach is the fact that the number of independent trials can no longer be increased by using many operators to inspect the same specimens.

3.2 Deriving probabilities from in-service data

In principle, the standard methods for determining probabilities can still be used with in-service data, however there are several difficulties to be overcome. Firstly, defects are only identified when they have finally been detected, possibly after several inspections. The overall number of misses is therefore not known as it is not possible to distinguish between a miss and a correct identification of "good" structure unless subsequent inspections indicate that a defect was present. The probability of detection must therefore be assessed from only the fraction of the defect population which has been detected. Secondly, if a complete reliability curve is required showing the detectability over a range of sizes during defect growth, it will be necessary to provide measurements of defect size. Finally, the provision of a complete inspection and service history for the components being inspected will be required. While this could be ensured for future systems the information may not be traceable with current operating practices.

At first sight, the first of the problems noted above appears the most important, namely that the total number of misses is never known and the probabilities are assessed on an unknown fraction of the defect population. In practice, this is unlikely to cause any significant problems. If the inspections are being carried out at a sufficient frequency to achieve a high safety level, then the standard methods of analysis can be used without accounting explicitly for the unknown additional misses as these will represent a negligible additional number. Since all defects are eventually detected, the probability of detection can be estimated from,

$$< n_e >= 1 / p_m$$

where $<n_e>$ is the average number of inspections required to detect those defects which were detected.

If there is a substantial risk of missing a defect completely, the mean probability of detection can still be estimated from the detectable crack data, correcting the observed mean probability to account for the finite number of observations. If the defects are inspected a maximum of r times, then the appropriate expression for $<n_e>$ is

$$< n_e >= 1/p_m + r\, q_m^{\,r}/(1 - q_m^{\,r})$$

where $q_m = (1 - p_m)$.

3.3 Length estimation by crack growth backprojection

A more serious, systematic problem for deriving reliability estimates from in-service inspection data is the lack of defect size information. In a typical, artificial POD verification exercise the defects are inspected in only one state and they are then measured accurately by destructive methods at the conclusion of the study. In the case of deriving reliability estimates from real data, only the final defect size can be known.

It was suggested that the solution to this difficulty is to use fracture mechanics to derive estimates of the crack length at previous inspections. The principle is illustrated in figure 3 for a typical fatigue crack. In this example it is assumed that the crack is found on the third inspection. The time of previous inspections where the defect has not been found are also known. A crack growth curve is used to back-project the crack growth from the known length where it was finally detected resulting in three data points, two misses and one hit at the appropriate crack lengths.

The crucial questions which arise in any attempt to use this approach are whether the random nature of crack growth will invalidate the procedure, what systematic errors will arise from using an incorrect crack growth curve to estimate the previous crack lengths and whether sufficient data is likely to be generated to allow useful reliability estimates to be obtained. Some guide to these can be obtained by simulations.

3.4 Simulation of reliability estimation from in-service inspection data

A simple model was used to simulate the crack growth and reliability curve generation problem. The crack growth was represented by a random process. In each period between inspections, the crack length was increased by a random increment proportional to the current crack length, generating a simple exponential mean growth curve. Inspections were also represented by a random process. The true probability of detection was defined for up to ten crack lengths. At each inspection, the current crack length and the inspection result were recorded. It was assumed that the series of inspections ceased whenever the crack was detected.

In order to generate reliability curves, a table of hits and misses in each crack length are required. Several assumptions were used to generate this information. In order to determine the best possible estimates, limited only by the size of the dataset, the actual crack lengths at each inspection were noted. These lengths would, of course, be unknown in a real situation. The most conservative assumption for crack sizes was to assume that no growth took place and the final crack length could be used for all of the inspections for a given crack. A back projection algorithm was used to simulate the process which would have to be followed in the real case. The back-projection algorithm assumed that the crack growth was deterministic following an average growth curve. Errors in estimating this growth rate were investigated by using several assumed rates in addition to the correct underlying rate which could only be estimated by fracture mechanics in a real situation.

The results of these assumptions are shown in Figure 4. The first row histogram shows the distribution of sizes at which hits were obtained. The second row shows the actual sizes of the cracks at all inspections, the information which we are trying to reconstruct. The third row shows the result of the conservative assumption that the cracks do not grow significantly and hence a_{last}, the length at the last inspection can be used to define the crack at all inspections. The distribution

obtained is similar to the distribution of crack lengths at which hits were obtained, but clearly it is very different to the actual distribution of crack lengths. The best crack length distribution obtainable from back-projection is shown in the fourth row, labelled "BACKPR". This was obtained using the correct mean growth curve. It is a good approximation to the actual crack sizes despite the random nature of the simulated crack growth. The final two rows show the results of underestimating and overestimating the crack growth rate by 50%. Underestimating the growth rate gives a conservative picture by overestimating crack lengths, while overestimating the growth rate leads to an underestimate of the defect population in all but the smallest size range.

Several reliability curves were generated from the data obtained. In order to estimate the best probability of detection estimates which could be obtained for p_m and p_α , these quantities were calculated using the actual crack lengths at each inspection using the 95% confidence level for p_α. The effects of random crack growth were investigated by calculating the reliability curve using the deterministic back-projection model and comparing with the curve calculated from the actual lengths. The effects of over and underestimation of the growth rate were assessed by using various assumed rates for back-projection. Figure 5 shows the reliability curves obtained from the defect populations illustrated in Figure 4.

It can be seen that the mean probability curve gives a fairly good approximation to p_t. The p_α curve obtained by back-projection is fairly close to the p_α curve calculated from the exact defect sizes, although for this simulation it seems to underestimate the reliability. This suggests that the method should be able to generate reasonably accurate reliability curves from real inspection data, provided the crack growth rate can be estimated accurately.

The p_m and p_α curves estimated from the constant defect size assumption can be seen to be completely the wrong shape reflecting the lack of realism in this assumption.

The p_α curves obtained from the final two data sets, using low and high back-projection rates respectively, show the expected systematic errors. Using too low a growth rate underestimates the probability of detection for most of the crack size ranges. Only at the lower end of the size range does the method overestimate the reliability as it fails to project all of the cracks down to these starting sizes. The opposite is true of the data using too high a growth rate. This assumption underestimates the number of trials which miss large cracks causing an overestimate of the detection probabilities.

Although these simulation results are encouraging there are two important caveats. The various p_α curves obtained are still conservative estimates of the probability of detection. Despite the very large number of defects used in the simulation, even the best p_α curve, that calculated from the actual crack lengths, only just reached the standard 90% level. The number of defects which can be expected to be seen in service is likely to be much lower than the 1000 used in the simulation. This will increase the separation between the p_t and p_α curves which have only just reached the standard value of 0.90. The other concern is the effect of the crack growth rate or equivalently the inspection interval. At slow growth rates many of the defects may be detected at relatively small defect sizes resulting in a small number of large defects in the larger sizes, with consequent low confidence bounds.

3.5 "Improved" estimates for the POD at large crack sizes

In the simulation described above, the crack growth rate was chosen to be approximately one crack length per inspection period. This allowed, on average, the appropriate three

inspections in the range where p_t is at least 0.90. In simulations where this fairly rapid, but random, growth takes place the reliability curves obtained were easily calculated from the large amount of data available. A significant number of defects were however missed completely due to fluctuations in the crack growth process rather than deficiencies in the inspection probability. A slower value of crack growth, averaging only half of the current size per inspection interval was also simulated. The results are shown in figure 6.

In this simulation, again using 1000 defects but with the crack growth reduced to 50% of the previous value, 75% of the defects were detected at the smaller defect size ranges where p_t is less than 0.90. This results in small defect numbers being available to determine p_α values. All of the p_α curves calculated for these larger defects fall below the desired 0.90 (except the overestimate arising from backprojection with too high a growth rate) and the curves actually fall away for the largest groups as there are few defects in these size ranges.

There are several standard procedures to increasing the estimates for p_α particularly toward the high end of the reliability curve. The principal methods are either to deliberately choose crack size intervals which lead to the highest values for p_α or to fit a parametric curve of some suitable monotonically increasing function through the mean probabilities at all crack sizes and calculate the α confidence limits on the parameters[10,11]. The former procedure, known as the Optimised Probability Method or OPM, introduces an unknown statistical bias into the calculations while the latter, used frequently during USAF studies of reliability, has the disadvantage that the results depend on making a good choice for the empirical curve. A variation on the USAF procedures is to use the inspection results for each inspection independently. The data to be fitted thus consists of a set of 1s and 0s. A suitable interpolating curve is then estimated by the maximum likelihood technique. This approach has been used in various studies and was most recently recommended by the FAA. These procedures were tested and reviewed extensively by Berens and Hovey[11] who favoured the curve fitting approach. They noted, however, that the method could not be used to deduce the crack size for which p_α first exceeded 0.90 as the results which it led to for this estimate were strongly affected by sampling errors and hence irreproducible. Either of these methods can be applied to the simulated data to "clean up" the reliability curves where lack of data reduced the values for p_α at large crack lengths.

The effect can be entirely removed by using the optimised probability method. Probability curves generated by applying the OPM technique to the slow crack growth data, i.e. the data illustrated in figure 6, are shown in figure 7 .

The reliability curves obtained have been forced into the expected monotonic increase with defect size. There is still a significant separation between the actual performance of the technique shown by the p_t and p_m curves and the 95% confidence p_α curves. The "correct" p_α curves corresponding to the actual defect sizes and the back-projected lengths at the average growth rates exceed 0.90 for only the last two size ranges where p_t has values of 0.95 and above. For size range 6, for example, the true probability is 0.92, however the corresponding estimates for p_α are 0.86 and 0.80. The vast majority of the defects, around 96%, were detected in the simulation in size ranges 1 to 6 leaving only 39 to be detected in the higher ranges.

The alternative curve fitting approaches to POD estimation do not really apply to analysing small data sets, particularly where there are few defects in the higher size (and probability) ranges. There is a danger that imposing a monotonically increasing curve

which asymptotically approaches unity will substantially overestimate the mean probability and 95% confidence bound for these large defects.

The 1000 defects considered in the above simulation studies are probably a luxury when compared to the defect numbers which can be expected in a study of real inspection data. The original Canadian pilot study[9] on cracks in the CF104 Starfighter was restricted to a data set of only 39 cracks. The above comments, particularly the concerns about the innate conservatism and lack of large crack data apply to smaller samples but obviously even more so. The results of 100 defect simulations at the faster and slow growth rate are shown below in Figures 9 and 10 to illustrate this.

The estimated numbers of trials in the larger crack size ranges, where p_t is known to be 0.90 or above totalled only 26, insufficient to verify a p_α of 0.90 at 95% confidence even if all trials were to be successful.

The reliability curves obtained in the standard crack growth model resemble those for the 1000 defect simulation, but due to the smaller numbers involved the confidence limits are considerably lower than in the larger simulations. None of the p_α curves reaches values as high as 0.90.

The slow crack growth rate model shows even more pronounced small sample effects. Although all cracks in the sixth range and larger were actually detected, the p_α curves all peak at a lower size at values well below 0.90. It is clearly necessary to use the OPM or curve fitting approaches to obtain any useful estimates for the large cracks.

The results of applying the OPM approach are shown in figure 11. In this particular simulation there is some discrepancy between the actual and back-projected crack length calculations which attain limiting values for p_α of 0.84 and 0.73 respectively, both well below the desired 0.90. Also, in this particular simulation, the overestimated growth rate back projection curve almost coincides with the "correct" p_α curve calculated from the actual crack lengths.

4. Alternative methods of analysis

4.1 Small samples from in-service data

The simulation exercises described above suggest that analysis of in-service inspection results can be used to estimate the inspection reliability. Use of real inspection data will overcome one of the principal objections to the use of reliability data based on artificial trials, the lack of realism particularly in human factors for an artificial experiment. The principal reservations appear to be whether the defect population can be estimated with sufficient accuracy and whether there will ever be sufficient data available to confirm the relative high and inflexible probability of detection standard currently required for airworthiness purposes.

If the NDE inspection results are likely to be insufficient to validate the canonical 90% POD at 95% confidence requirement it is necessary to assess whether there is a better way of measuring and reporting NDE reliability. Viewed from the NDE perspective, the need is for a statistical method of analysing the data which will most efficiently make use of whatever data can be collected to predict the probable outcome of future inspections.

The simplest case can be thought of as the task of predicting the probability of missing a defect during the number of inspections, typically three or so, which will be carried out in service. This assumes that the probability of detection is constant, as is assumed in the current methodology. There are various approaches to statistical inference as this form of prediction is known. Such approaches usually take the form of trying to predict as

accurately as possible the probability of an expected outcome. The standard method of analysing NDT reliability, establishing a lower bound and then using this to estimate the probability of missing a defect three times, say, is unusually conservative.

4.2 Inefficient predictions from current methodology

The extent of the inefficiency in the estimates incorporated in the standard methodology can be illustrated by estimating the probability of missing a defect three times, given that the required p_α of 0.9 at 95% has been verified. The conservative estimate is simply obtained by assuming that $p_t = p_\alpha$ in which case the probability is simply 0.001. In reality, in order to verify the p_α value, the actual p_t value for the technique must be higher than 0.90, somewhere closer to p_m. Using p_m to estimate the outcome of the three inspections leads to the values in table 1 where the results of the initial verification exercise are given, together with the resulting p_m and the most likely prediction for probability of three misses.

Initial experiment		Probabilities		p_α = 0.90
			Prob of 3 misses	
Hits	Trials	p_m	$(1-p_m)^3$	$(1-p_\alpha)^3$
29	29	1	0	0.001
45	46	0.978	1.03E-05	0.001
59	61	0.967	3.52E-05	0.001
73	76	0.961	6.15E-05	0.001
85	89	0.955	9.08E-05	0.001
98	103	0.951	0.000114	0.001
122	129	0.946	0.00016	0.001
157	167	0.940	0.000215	0.001

Table 1, Estimated safety level, i.e. probability of three successive misses, after verifying a p_α of 0.90 at 95% confidence.

It can be seen that the likely performance of the technique is very much better, possibly an order of magnitude better than the conservative estimate predicts. This degree of conservatism is acceptable if sufficient information is available to verify the high p_α value, however it is a luxury if it is unrealistic to expect the limited data available to provide such high estimates.

If NDT methods are capable of achieving a reliability of above 90% it may be more realistic to set a lower value for the required p_α value. If we consider the trials which would be required to establish the lower p_α value of 0.85 rather than 0.90, again at 95% confidence, we find the following results shown in table 2.

Initial experiment		Probabilities		p_α = 0.85
			Prob of 3 misses	
Hits	Trials	p_m	$(1-p_m)^3$	$(1-p_\alpha)^3$
19	19	1	0	0.0034
28	29	0.966	4.1E-05	0.0034
38	40	0.950	0.00013	0.0034
47	50	0.940	0.00022	0.0034
55	59	0.932	0.00031	0.0034
63	68	0.926	0.00040	0.0034

Table 2, Estimated safety level, i.e. probability of three successive misses, after verifying a p_α of 0.85 at 95% confidence.

Again it can be seen that the conservative estimate is an order of magnitude or more worse than the expected performance. although the technique would be considered inadequate to meet the standard requirement, the estimate of its performance based on p_m the best estimate of the true capability of the technique is still significantly better by a factor of at least two than the required minimum performance. A method of statistical inference which preserved more of the information from the verification experiment, or equivalently from the in-service experience so far, would perhaps be able to give the necessary safety level prediction of 1 missed defect in 1000, for techniques where the available data would not confirm the standard p_α value of 0.90 at 95% confidence.

4.3 Bayesian inference for NDE assessment

There are various methods of predicting the probability or likelihood of an outcome based on an initial experiment. The most straightforward are based on the use of a contingency table and a standard statistical test

such as the χ^2 or Fisher's likelihood test. These approaches allow the probability of missing a defect three times after the initial experimental result to be deduced directly without recourse to calculating an intermediate POD for a single trial.

A more elegant method can be based on Bayesian inference[12]. In the Bayesian approach, the degree of confidence in a particular outcome before an experiment is expressed as a "prior" distribution of probabilities. In applying the approach to NDT reliability assessment, the prior distribution is chosen as the level of confidence in achieving given values for the probability of detection. An initial experiment is then carried out. The outcome of the initial experiment is used to update the prior distribution, producing a "posterior" distribution reflecting the revised degree of confidence in the possible outcomes as a result of including the additional information which has been obtained. Bayesian confidence levels and intervals can be estimated from the posterior distribution. Finally the Bayesian analysis can be used to produce a third distribution, the "predictive" distribution, which is calculated directly from the posterior distribution. This gives the probability of any outcome in a subsequent experiment given the initial level of knowledge in the prior distribution and the additional information from the initial experiment.

A useful concept in Bayesian analysis is the use of conjugate pairs of distributions. The results of the experiments can be described by one type of distribution, in this case the binomial distribution. If a prior distribution can be chosen from a family of distributions so that the posterior distribution calculated from the experiment is from the same family as the prior distribution, then the two distribution types are said to be conjugate. Since the prior and posterior are of the same type, it follows that any further experiments can be used to generate a further posterior distribution incorporating all of the experimental information which will again belong to the same family of distributions. In the case of the binomial distribution $p(h,n,p_t)$, it is known that the conjugate distribution is the Beta distribution $Be(\gamma,\eta,p)$ where γ and η are constants. The predictive distribution formed from the Beta distribution is called the Beta-Binomial distribution, $BeBi(h_2,n_2,\gamma,\eta)$ where h_2 and n_2 are the assumed hits and trials in the subsequent experiment. Full details are given in ref 12.

The prescription for analysing reliability experiments in this formalism is then to start with a prior distribution from the Beta family. An initial experiment or a series of inspections in service will provide a known number of hits and misses which can be used to update the prior. It can be shown that if the prior is $Be(\gamma,\eta,p)$ and a binomial experiment has resulted in h hits and n - h misses, then the resulting posterior distribution is $Be(\gamma+h,\eta+n-h,p)$ and the predictive distribution is $BeBi(h_2,n_2,\gamma+h,\eta+n-h)$. The results of subsequent experiments or periods of inspections in service can naturally be built into the posterior distribution by using the total numbers of hits and misses to date.

The process can be illustrated by simulating a reliability verification experiment carried out in small sets of trials. In the example below it is assumed that the underlying probability of detection, the true probability p_t is 0.92. A total of 45 inspections has been carried out in groups of 5 inspections. An initial prior distribution has been chosen with $\gamma = \eta = 1$ which gives a uniform distribution indicating that no information on the reliability of the technique is available. Figure 12 shows the posterior distributions after each set of 5 trials.

The actual simulation depicted resulted in 4 misses in the 45 trials for an average probability $p_m = 0.911$. The evolution of the posterior distribution shows that it is quite broad after the initial sets of trials, but rapidly

becomes more peaked around the mean probability value. The confidence level for any value of the probability of detection p can be obtained directly from the posterior distribution. For comparison, the evolution of the estimates for the mean and 95% p_α are shown in Figure 13. For such a small number of trials the 95% p_α is well below 0.90.

The safety level can be calculated from these probabilities p_m and p_α and from the Bayesian predictive Beta-Binomial distribution, again after each 5 trials. It is assumed that three inspections will be carried out in service on the defects, hence the appropriate expressions are

Binomial;

$$p(0, 3, p_{\alpha/m}) = (1 - p_{\alpha/m})^3$$

Bayesian;

$$p(0,3) = \mathrm{BeBi}(0, 3, 1+h, 1+n-h)$$

The three probabilities are shown in figure 14. It can be seen that the true safety level does indeed attain the desired 0.001. The Bayesian estimate of the safety level is conservative, however it is significantly closer to the real value than the classical estimate from the 95% lower bound on the POD.

The greater efficiency in translating the full available information on NDT reliability into a direct estimate of the safety level which can be expected offers the possibility that useful reliability statistics and safety level estimates can be generated from substantially less data than would be required for the standard POD analysis. This approach requires further investigation.

5. Conclusions

The standard analysis of NDT reliability using artificial experiments to verify a lower bound to the probability of detection of 0.90 at 95% confidence is difficult to apply to airframe structural inspections due to the difficulty or cost of providing a large number of realistic trials. The built-in conservatism may be excessive making it impossible to verify the capabilities of adequate inspection techniques or leading to unrealistic estimates for the required frequency of inspection.

The use of in-service inspection data to assess inspection reliability may overcome the lack of realism in reliability assessment exercises. The data collected may be limited to small numbers of defects, especially in the larger defect sizes. Alternative analysis methods may be necessary to demonstrate acceptable levels of safety for airworthiness purposes without invoking the standard probability of detection of 0.90 at 95% confidence.

More efficient statistical methods can demonstrate higher safety levels than the standard analysis. This may not be necessary for situations where there is adequate reliability data to use the standard methods. It may be crucial to reliance on NDT where verification of high reliability is limited by available data.

One approach based on Bayesian inference was described and shown to be able to give useful quantitative estimates for safety levels on very limited data. Further analysis of this, or other approaches which make the best use of limited data, should be undertaken to provide a more flexible alternative to the standard methodology.

6. References

1) USAF Military Specification, airplane damage tolerance requirements. MIL-A-83444 (1974)

2) "Assessment and Demonstration of the Capabilities of NDI Processes, Equipment and Personnel" Rummel W D in "Impact of emerging NDE-NDI methods on Aircraft Design, Manufacture and Maintenance" AGARD-CP-462 (1989)

3) "Reliability of Flaw Detection by Nondestructive Inspection" Packman P F, Klima S J, Davies R L, Malpani J, Moyzis J, Walker W, Yee B G W, and Johnson D P in AMS Metals Handbook Vol 11 NDI and Quality Control.

4) "Recommended Practice for a Demonstration of Nondestructive Evaluation (NDE) Reliability on Aircraft Production Parts" Rummel W D Mat Eval **40** 922 (1982)

5) " Bruce D A and Stone D E W in "Reliability in Non-Destructive Testing" ed Brook and Hanstead (Pergamon, Oxford) p45 (1988)

6) "Statistical Theory with Engineering Applications" Halb A Wiley New York p726 (1952)

7) "Inspection Reliability" Bruce D A in "Impact of Emerging NDE-NDI Methods on Aircraft Design, Manufacture and Maintenance" AGARD-CP-462 (1989)

8) "Non-Destructive Inspection and the implementation of a damage tolerant design philosophy" Stone D E W Tech Memo Structures 982 (1981)

9) "Development of Non Destructive Inspection Probability of Detection Curves

Using Field Data" Simpson D L Laboratory Technicval Report LTR - ST - 1285 NRC Ottawa Canada 1981

10)"Reliability of Nondestructive Inspections" Lewis W H, Dodd B D, Sproat W H and Hamilton J M Lockheed Georgia Company report SA-ALC/MME 76-6-38-1

11)"Evaluation of NDE Reliability Characterisation" Berens A P and Hovey P W Technical report AFWAL-TR-81-4160 (1981)

12)"Statistical Prediction Analysis" Aitchison J and Dunsmore I R (Cambridge University Press) (1975)

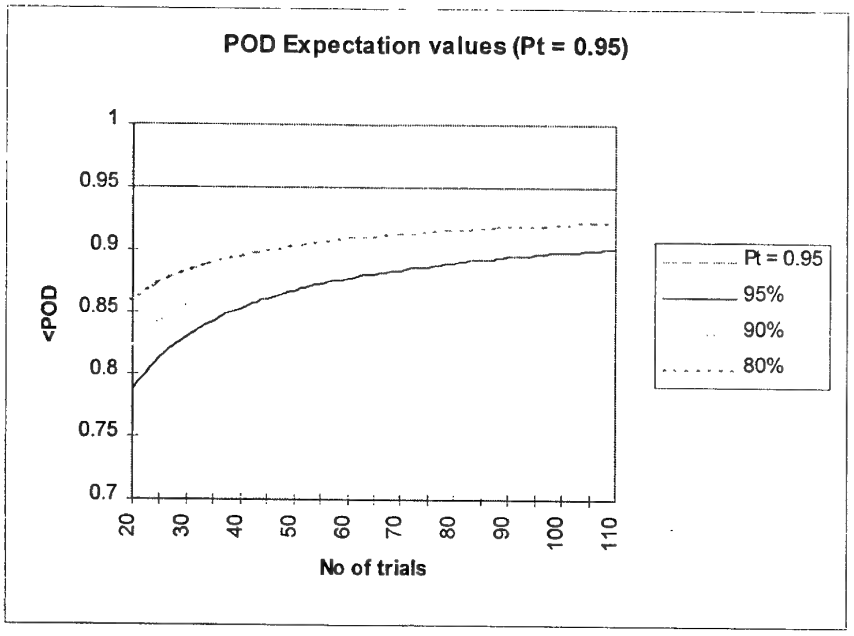


Figure 1, Expectation values for p_α for confidence levels of 80, 90 and 95% for a highly reliable inspection process where p_t is 0.95.

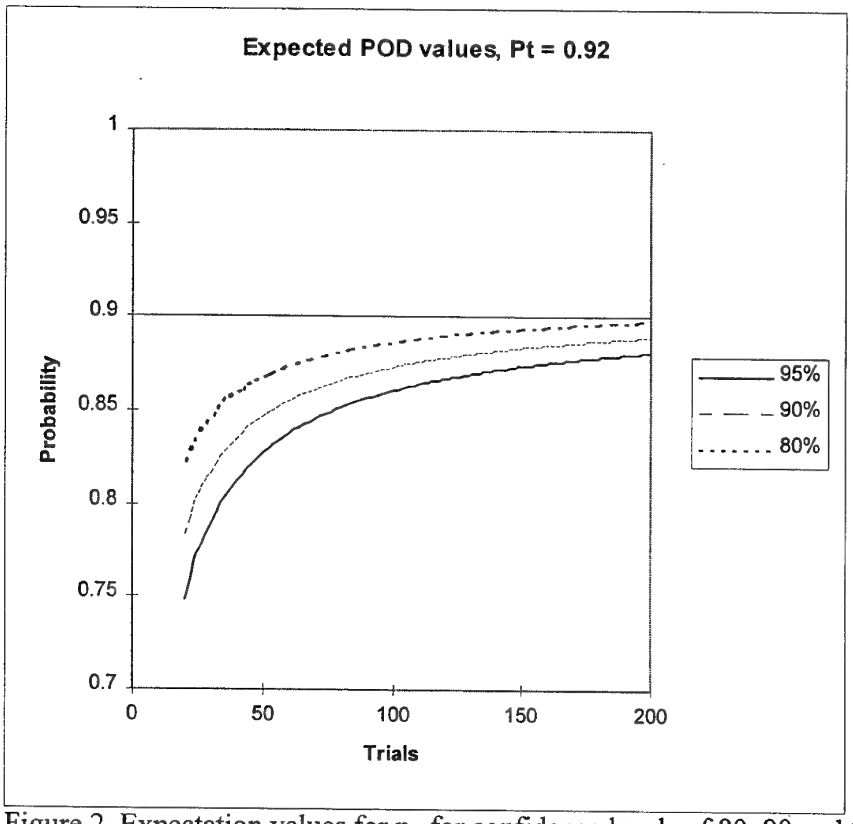


Figure 2, Expectation values for p_α for confidence levels of 80, 90 and 95% for an inspection process where p_t is 0.92.

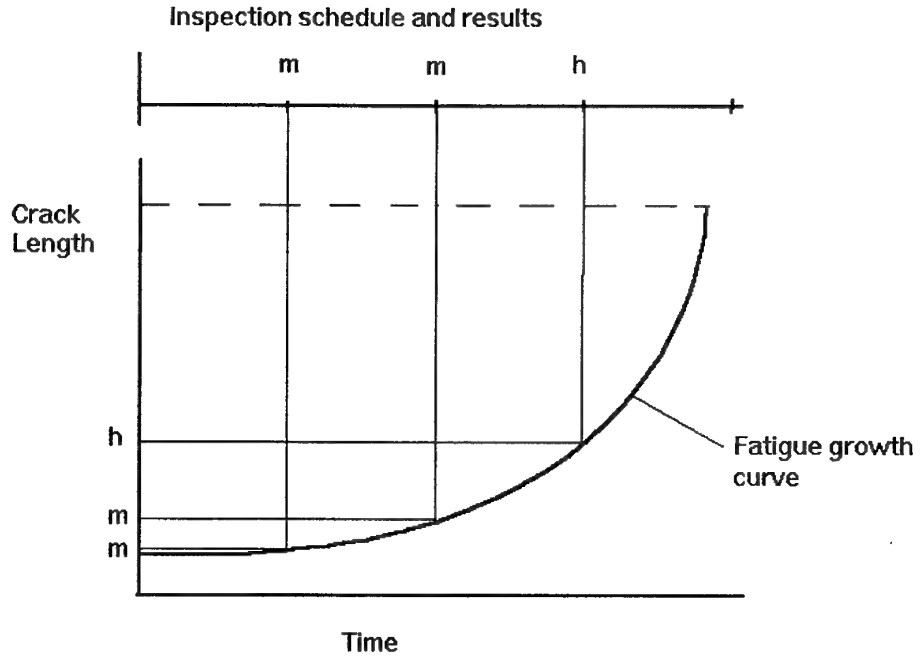


Figure 3, generation of reliability data from a single fatigue crack.

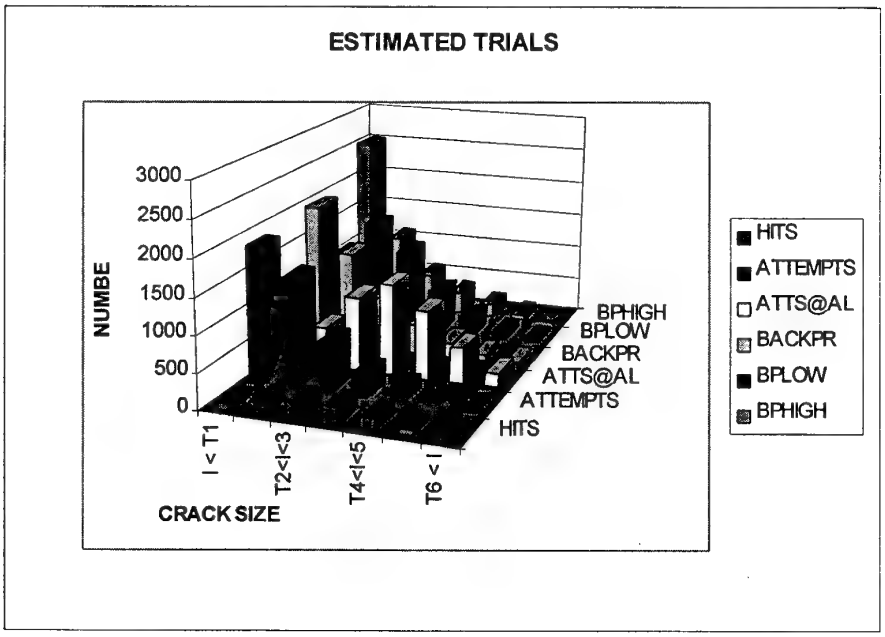


Figure 4, Estimated defect size populations for a 1000 defect simulation

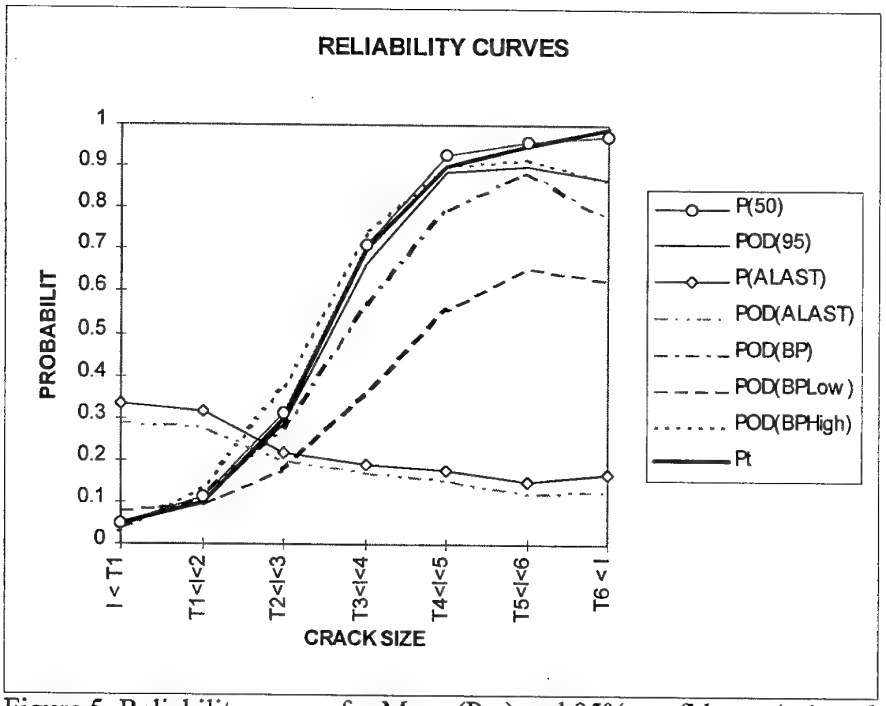


Figure 5, Reliability curves for Mean (Pm) and 95% confidence (pa) probabilities of detection estimates compared to the true underlying reliability curve Pt.

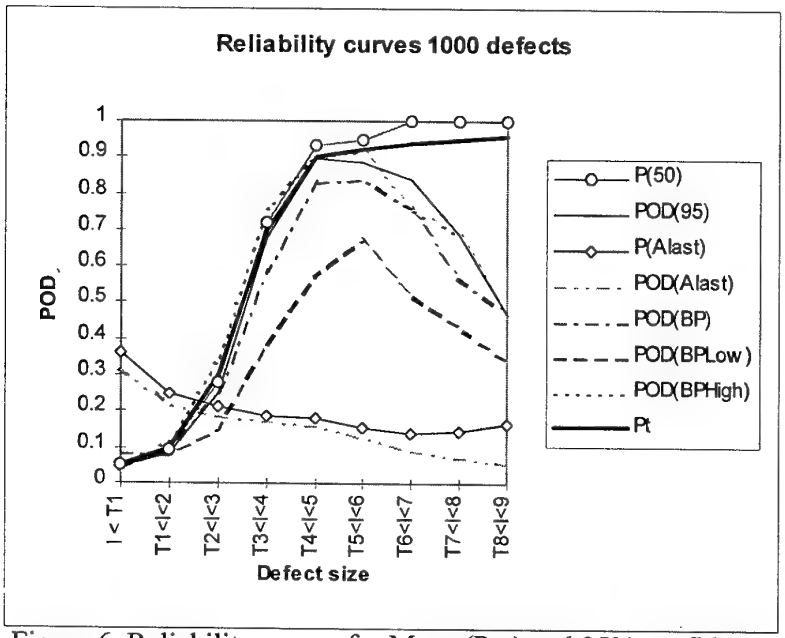


Figure 6, Reliability curves for Mean (Pm) and 95% confidence (pa) probabilities of detection estimates compared to the true underlying reliability curve Pt, assuming a slow crack growth rate.

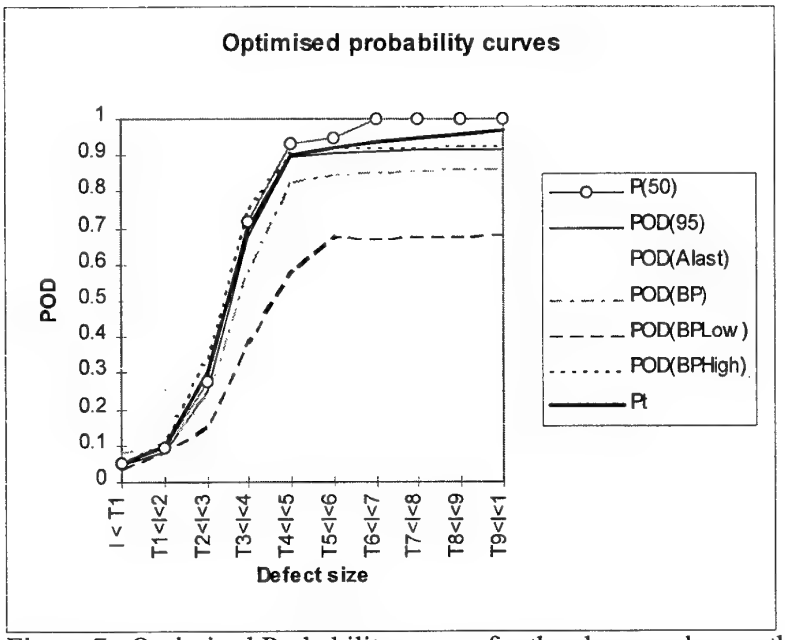


Figure 7, Optimised Probability curves for the slow crack growth rate simulation

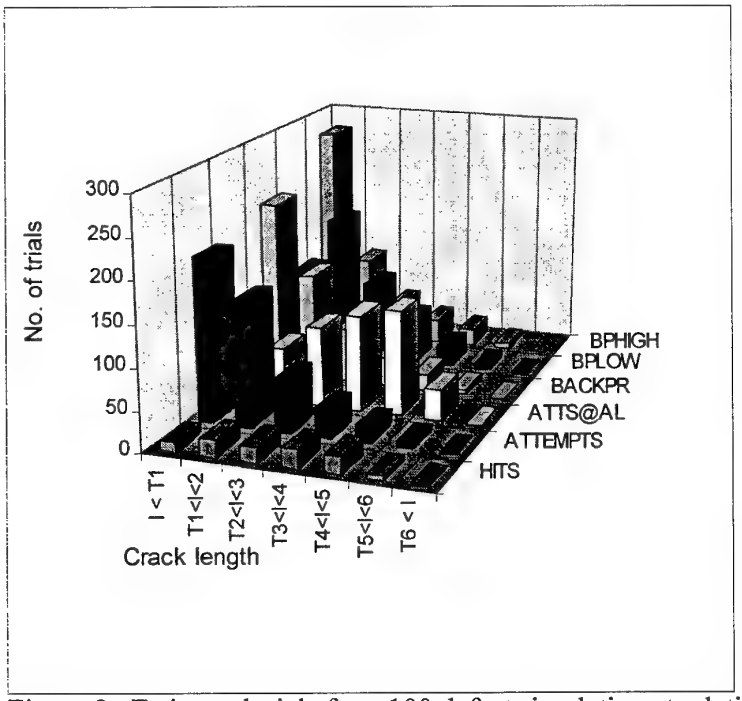


Figure 8, Estimated trials for a 100 defect simulation at relatively fast crack growth. Note the small number of trials at size ranges 5 and above.

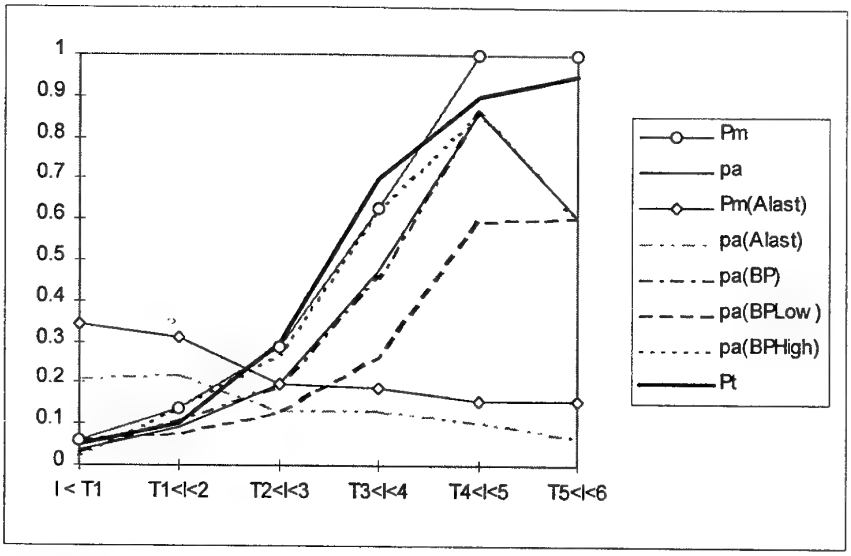


Figure 9, Reliability curves for Mean (Pm) and 95% confidence (pa) probabilities of detection estimates for 100 defects compared to the true underlying reliability curve Pt.

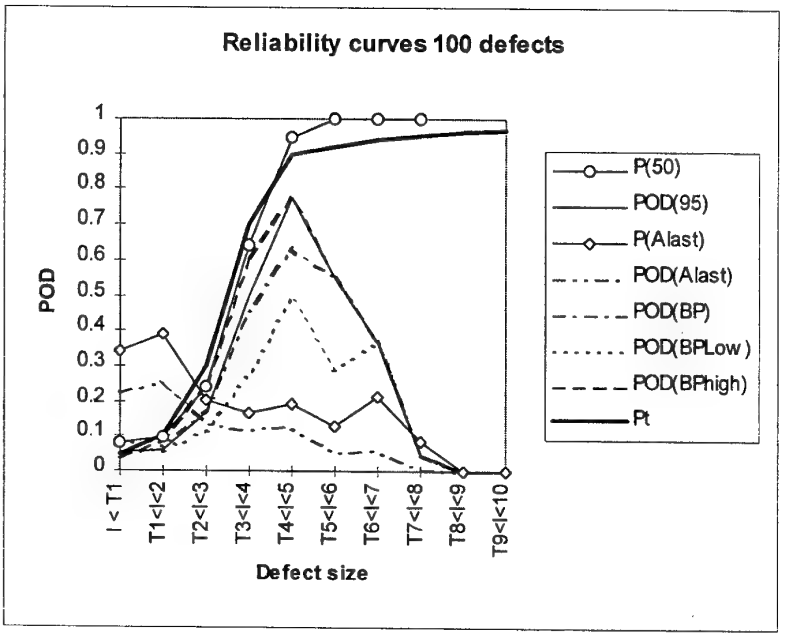


Figure 10, Reliability curves for Mean (Pm) and 95% confidence (pa) probabilities of detection estimates for 100 defects, assuming a slow growth rate, compared to the true underlying reliability curve Pt.

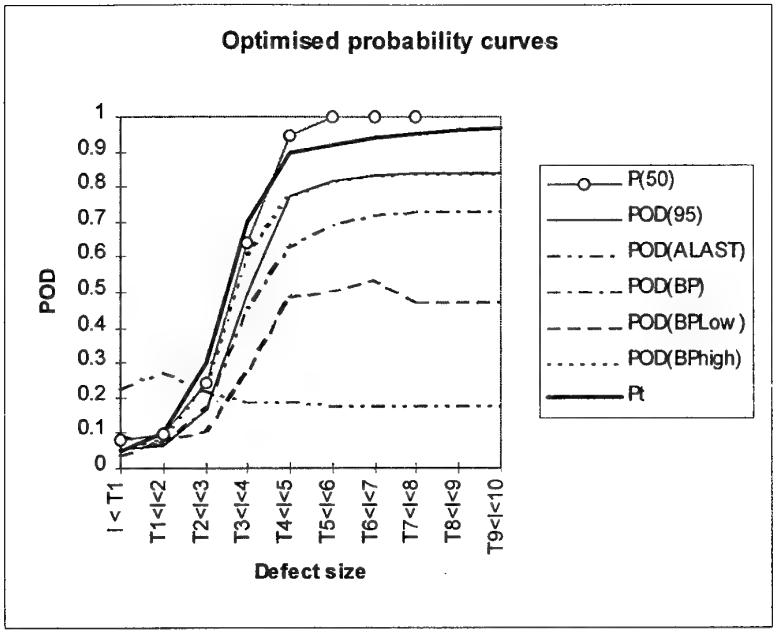


Figure 11, Optimised Probability curves for the slow crack growth rate simulation of 100 defects

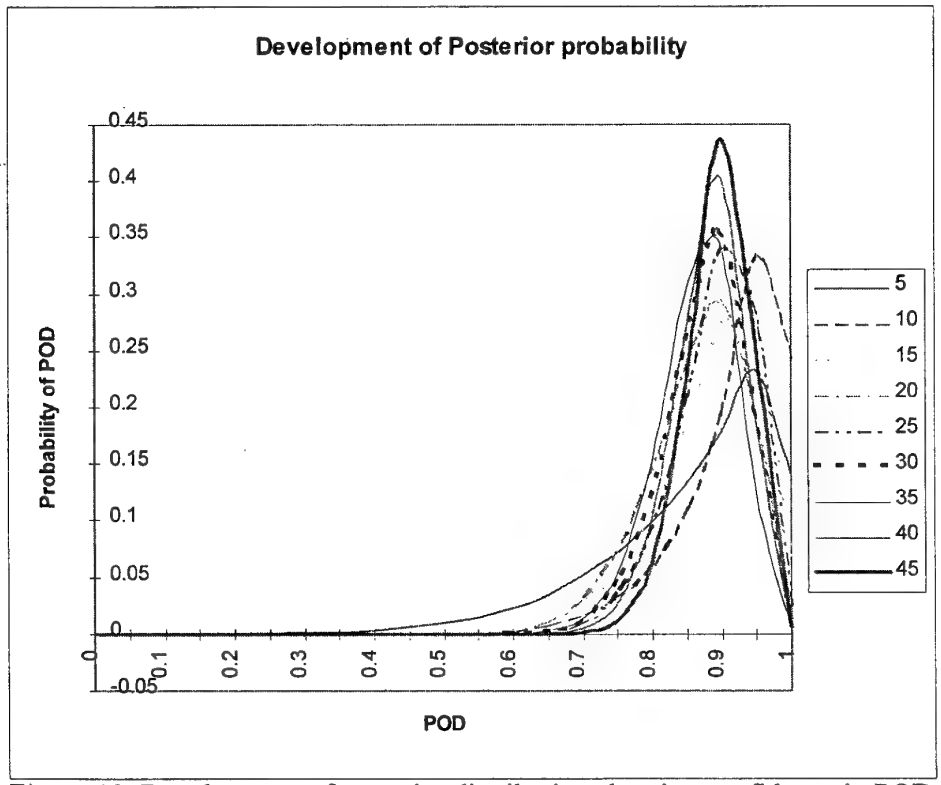


Figure 12, Development of posterior distribution showing confidence in POD

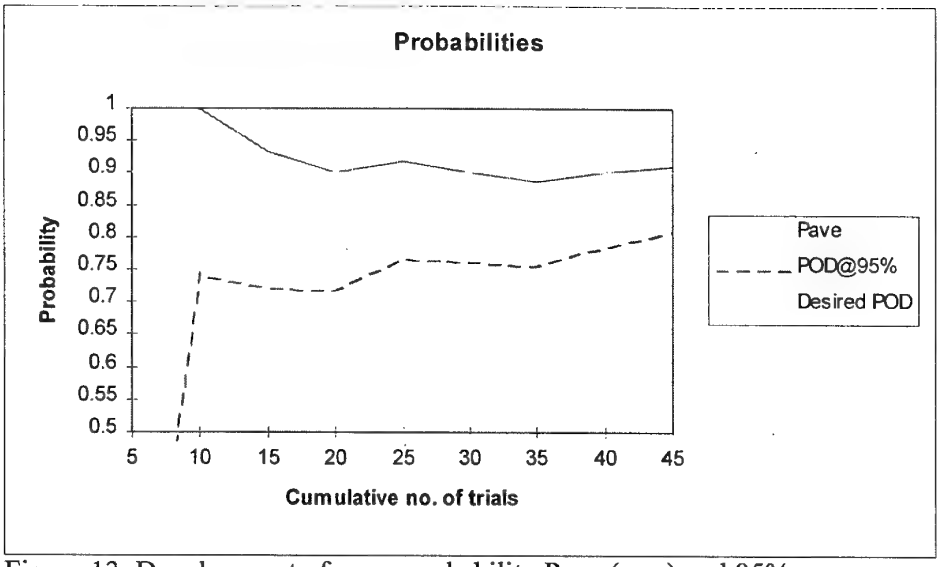


Figure 13, Development of mean probability Pave (=p_m) and 95% p_α

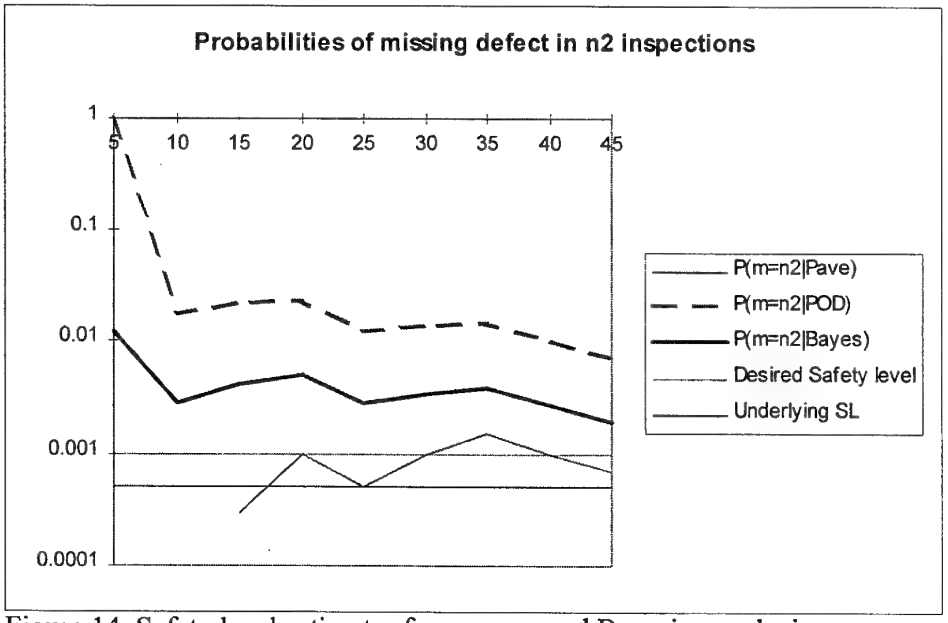


Figure 14, Safety level estimates from p_m, p_α and Bayesian analysis.

ROYAL AIR FORCE IN-SERVICE APPROACH TO AIRFRAME INSPECTION RELIABILITY UNDER FIELD/DEPOT CONDITIONS

Flt Lt S J Beverley
Non-Destructive Testing Squadron
Royal Air Force St Athan
Barry
Vale of Glamorgan CF62 4WA, UK

SUMMARY

The reliability of an NDT technique is directly attributable to 3 factors: the chosen methodology; the quality of technique that the NDT technician is applying; and most importantly, the technicians' capability. The RAF infrastructure optimises each of these 3 elements to ensure that any inspection can be reliably carried out by any NDT technician. NDT techniques are based on practical research and are validated by selected personnel, who have numerous years of operational experience, using all NDT methods. Equally adept, highly motivated technicians then apply the techniques. Hence, confidence in the quality and repeatability of any NDT technique within the RAF is extremely high.

The paper will describe the organisation and functions of the RAF NDT Squadron. In-Service equipment and NDT Squadron capability will be reviewed, followed by a look at evaluation and procurement of new equipment, as a function of tasks and technological advancement. The merits of assessing the probability of detection (POD) within the NDT Squadron, using a sterile environment for data capture will then be discussed. The paper will move on to discuss how technique validation is the key to maintaining NDT reliability, and that metallurgical data on the minimum fault size to be detected is essential when selecting the correct methodology.

A number of case studies will be presented to highlight some of the more challenging techniques that have been developed and how the various sections interact to provide an effective NDT solution.

1 INTRODUCTION

It is fundamentally accepted that the POD of a fault is dependent on 3 key factors: the methodology; the equipment capability; and finally, the operator or human factors. However, one further factor, the 'operational requirement', has an over-riding effect on the development and application of any technique, and can ultimately have a more serious effect on the detection capability of any technique. Hence, the role that NDT must play in maintaining operational effectiveness is to provide the highest degree of confidence that a fault or defect has not been missed. Furthermore, to optimise the technique's reliability, the parameters of a specific inspection technique must be such that they include a margin of safety, thereby reducing the human and operational factors on the final outcome.

In order to achieve these aims, a solid organisational infrastructure is required to support the technicians working in industrial environments. The current organisational structure of the RAF Non-Destructive Testing Squadron has evolved over the past 40-years into its current format. The interactions between the various flights enable the Squadron to react rapidly to any situation; NDT R&D technicians supporting Technical Authors who are responsible for the production of techniques issued to the Regional NDT Teams, who functionally undertake the inspection.

2 NDT SQUADRON ORGANISATION

NDT Squadron contains 5 major elements: the RAF School of NDT; the Regional NDT Teams; Technique Development Section; Equipment Evaluation Flight; and finally, the Repair and Calibration Flight. The Squadron is also accredited to ISO 9001, demonstrating a commitment to quality assurance practices.

Selection of NDT Technicians. All RAF NDT Technicians must have a minimum rank of Sergeant before they can apply for specialist duties with the Squadron. Invariably, this means that they have gained a minimum of 10 years trade experience in the airframe, propulsion, or aircraft electrical trades. Additionally, they will have demonstrated a good standard of trade competence and have a good aptitude for both Mathematics and English. Selection is then made by interview and the candidate must demonstrate a high degree of motivation for employment in the specialisation. Following selection the technician is then enrolled with the RAF School of NDT for his professional training.

The RAF School of NDT. The RAF School of NDT provides training to PCN Level II for all NDT technicians employed by the MoD. The school runs 4 principle courses: The NDT Technicians Course; The Technicians Update Course; The NDT Appreciation Course; and finally, the Care and Use of Remote Viewing Aid Equipment Course. In addition, the school is able to run specialist courses when training is required on new equipment types.

The NDT Technicians Course runs for 10 weeks and covers the 6 standard disciplines of PFD, Visual Aids,

Paper presented at the RTO AVT Workshop on "Airframe Inspection Reliability under Field/Depot Conditions", held in Brussels, Belgium, 13-14 May 1998, and published in RTO MP-10.

Magnetic Particle Inspections, Eddy Current Inspections, Ultrasonics, and Radiography. As part of the course, the students are also given a comprehensive brief on the nature and origin of faults, and Radiation Safety. The course is tailored to meet the specific requirements of aerospace inspections and all examinations reflect the standard operating procedures that are used within the RAF. Following successful completion of the academic phase of the course, students are allocated to one of the RAF Regional NDT Teams, where they spend a minimum of 6-months under supervision, before finally proving their competence in all disciplines. The school is accredited to PCN Level II for training, however, the examinations are generated in-house. The respective team managers are responsible for monitor technical performance standards.

On a 3-yearly cycle, all technicians return to the School to be re-examined in Radiation Safety and all 6-disciplines, thereby maintaining their core knowledge. In addition to the examinations, the technicians receive continuation training on the latest technological advances.

The NDT Appreciation Course is designed to provide personnel with an overview of the techniques and capabilities of the NDT Headquarters Staff and the Regional NDT Teams. The course is primarily aimed at those personnel employed as Support Authorities (SA) for specific aircraft types, as they are responsible for the development of maintenance inspection programmes.

Finally, the school runs a course on the care and use of remote viewing aid equipment. The course is designed to increase the awareness of personnel using the equipment and of the various systems' capabilities.

Regional NDT Teams. The Regional NDT Teams have the responsibility of supporting the front line squadrons for any NDT tasks that may be required. There are 8 teams based in the UK, plus one in Germany, based at RAF Bruggen. Additionally, NDT Squadron supports the Falkland Islands with one technician. There are 60 personnel serving with the Regional NDT Teams and, for UK tasks, their tasking is controlled by their respective team managers. However, should the need arise for personnel to support deployed operations, the NDT Control Officer, based with the NDT HQ, has functional command and he is able to provide support, on a case-by-case basis, within 24-hours of being tasked.

Technique Development Section. The Technique Development Section is responsible for producing and reviewing all inspection techniques applied to the majority of MoD aircraft. The personnel employed in the section have a minimum of 3 years field experience and they are responsible for a number of aircraft types. Invariably, personnel are allocated aircraft that they have a first hand NDT and structural knowledge of, thereby enabling them to provide expert advice to aircraft SA. The principle objective of any technique issued is that it should be reliable and repeatable, as any one of the Regional NDT personnel may have to apply the technique.

Tasking from an aircraft SA will specify the requirement to develop a technique, and the time-scale in which a response is required, which can be as little as 24 hours. They will also provide information supplied by metallurgists and stress office personnel on the fault size they wish to find, and on the maintenance frequency at which the technique will be implemented.

Equipment Evaluation Flight. The Equipment Evaluation Flight operates in support of both the Technique Development Section and the Regional NDT Teams. It is the responsibility of the Flight to maintain an 'intelligent customer' capability, by maintaining close working relationships with equipment manufacturers and research organisations. Hence, when unusual situations arise which require special-to-type probes or diverse applications, such as acoustic emission or laser shearography, the Flight has the capability to recommend the 'best practice' technique for a specific task. The technicians employed within the Flight have several years' experience with the regional NDT Teams before being specially selected for this employment. Each technician has a principal specialisation i.e. Ultrasonics, Eddy Current or Radiography, and 2 technicians are employed to manage the vast range of remote viewing aid equipment that is currently being used throughout the Royal Navy, Army and RAF.

Repair and Calibration Flight. The Repair and Calibration Flight provides the Squadron with a depot facility for the maintenance, storage and distribution of all NDT equipment. Furthermore, the Flight is able to incorporate both hard and software upgrades to our in-service equipment. In addition to the standard NDT equipment, they provide the same facility for the management of helicopter and engine vibration analysis equipment.

3 TECHNIQUE DEVELOPMENT

Selection of the inspection methodology is wholly reliant on the experience of the personnel working in the section. A high degree of confidence is entrusted in the personnel, as we have already established that they are high calibre personnel who have received professional training and have a number of years field experience before being posted into the job. Additionally, the development of any technique is subjected to several levels of scrutinization before it is finally issued.

Before a technique is formally issued it has to be validated against some known fault criteria. Due to the operational constraints under which the Squadron operates, it is impossible to undertake any formal POD analysis before a technique is issued. Therefore, in order

to provide a high degree of confidence that the technique will be both reliable and repeatable, every technique goes through a thorough validation process which involves simulating faults using EDM notches in components that model the structure found on the aircraft. The technique is then finally proved against a component with a known fault. Wherever possible, validation components will have several NDT methods applied to them in order to ascertain the exact dimensions of the known fault criteria.

Where inspection techniques are deemed to be of a non-standard nature then additional validation pieces are provided to the Regional NDT Teams as training samples. This enables the technicians to ensure that the equipment is correctly set-up and that he is able to resolve faults against the specified criteria.

With the Squadron being based at RAF St Athan, easy access to the majority of RAF aircraft is facilitated. Hence, we are able to develop the techniques in a real-time environment. Additionally, with the Unit having structure manufacturing and component overhaul facilities we are able to supply calibration samples that conform to the requirements of the techniques.

Categories of Technique. Techniques are either classified as Category A or B. Category A techniques involve the application of a methodology that can only be carried out by a Level II trained person, as there will be a requirement for some aspect of signal analysis to be performed. Hence, these are constrained to NDT technicians currently employed within NDT Squadron. Category B techniques are those which are delegated to non-specialist technicians who are required to use some aspect of lower-level NDT as part of their duties. Typically, these tasks include the use of Visual Aid equipment, PFD techniques or some automated process, such as the eddy-current inspection of wheels.

All Category B operators are trained by Regional NDT Team personnel with update training and re-certification being carried out on a 6-monthly basis. In addition, the technician must maintain a degree of currency in the application of the technique, within the 6-month period, in order to maintain his certification for the task.

4 EQUIPMENT EVALUATION

It is very easy to be sold an item of test equipment that is demonstrated against its optimum performance under laboratory conditions. However, when applied in the hangar environment, other criteria may affect the signal processing, or the logistics of applying the equipment to the inspection surface may make the system unusable. Hence, there are a number of objectives that must be attained when undertaking any evaluation of equipment, specifically:

a. To determine that the equipment has an improved capability over existing technology.

b. To establish that the equipment is user-friendly.

c. It satisfies the operational constraints for in-Service inspections.

With improving technology, the man-machine interface is becoming user-friendlier and the image analysis becoming simpler for the technician. However, the greater the extent of pre-processing that takes place often means that more detailed sensitivity tests against calibration panels must be undertaken. The design of test panels is critical to the selection process, as the realistic modelling of faults is essential to determining the true capability of a system.

Visual Aids Procurement Standards. As the majority of visual aid inspections are Category B techniques, we have to ensure that the equipment being used is capable of resolving faults to the specified standard. NDT Squadron has been proactive in raising the manufacturing standards of remote viewing aid equipment with the development of an optical test bench, which automatically analyses both the on-and-off axis light transmission through a simulated fault onto a CCD unit. The system is then able to determine the optical quality of a probe.

Digitization of Radiographs. The latest developments in radiographic processing involve laser scanning of film, either by digitizing conventional film or by using the photo luminescence of phosphor screens, to produce a digital radiographic image. In either case, laser scanning systems have been developed to scan the images at 12 bits, providing images with 4 096 grey scales. At best, the human eye is only capable of identifying 128 grey scales or a 7-bit image.

Obviously, there are several issues that require addressing when looking at either digital or phosphor screen radiographic processing, such as the scan resolution and the film sensitivity. Notwithstanding these issues, there is the potential in these systems to improve radiographic inspection standards using image manipulation techniques. Additionally, the inspectors' eyes suffer less strain viewing a monochrome monitor than he does a high intensity radiographic viewer, thereby reducing one potential human factors issue.

Area-Scanning Systems. With the development of IT systems for C-Scan mapping components, there has been a change in emphasis in the way NDT is applied at the service level. Previously, C-Scan mapping was constrained to laboratory conditions. However, systems have now been developed for field use, enabling graphical representations to be made of fault areas. Additionally, the use of these systems has changed the

emphasis of NDT applications in the service environment. We are now able to monitor specific areas and provide detailed maps of the defects.

There are numerous technologies that fall into the category of area scanning systems, ranging from conventional eddy current and ultrasonic systems interfacing with laptop PCs, to optical systems that utilise thermography or shearography. There are benefits to each of the systems; however, for the aerospace industry, the field application of some technologies is still a number of years away.

5 OPERATIONAL INFLUENCES ON NDT POD

The primary desire of any SA is to have any inspection carried out in-situ, rather than incurring the additional financial cost and maintenance penalty of having to remove a component. Invariably, at some stage of deep maintenance, any component can be inspected with an absolute certainty of determining if it is fault free. However, this fundamentally defeats the principal objective of NDT. Hence, the economics of scale have a major bearing on the inspection methodology used.

Once it is established that the methodology and technique have been optimised, and the personnel have been trained and qualified in the application of the particular technique, it would be foolish to presume that a technique would have a 100% POD. No matter how ideal the working environment is, some external factor will influence the final outcome of the inspection. Physiological and psychological studies have proven on a number of occasions that human factors will always degrade the final outcome of any repetitive task.

Human Factors. Repetitive tasks will inevitably be those where human factors will have the most significant effects. Therefore, it requires a highly motivated technician to apply the same tenacious approach to inspecting the 100th component as he did to the first component examined.

Where NDT requires signal interpretation, the technician may have to make qualitative decisions, and hence there will invariably be differences of opinion between technicians. Additionally, the interpretation of optical images is wholly dependent on the technicians' eyesight being able to resolve the fault standard. Large radiographic inspection tasks are a good example of the repetitive nature of some tasks. For example, technicians can expect to spend up to 5 working days viewing the radiographs required for Nimrod Major maintenance. Sufficient rest periods are essential to maintain not only the concentration of the technician, but also his eyes from becoming tired from the light emitted from the viewing screens. RAF technicians are only permitted to view radiographs for a continuous period of 20 minutes, after which they must take a 10-minute break from that working environment.

CASE STUDIES

The best means of highlighting some of the issues that face in-service inspection teams, and organisations such as NDT Squadron, that have a robust technical infra-structure designed to produce inspection techniques that are optimised for field conditions, is by looking at a number of case studies. Each takes a brief look at some of the issues that affect POD criteria. The last of the case studies will introduce some work that has recently commenced to evaluate a true POD for a technique that was influenced in its development by a number of factors.

Jaguar Frame 25 Inspection. Frame 25 on the Jaguar aircraft has been identified as suffering from stress-corrosion cracking. The area to be inspected is in the bore of the undercarriage mounting. An eddy current technique was developed using impedance-plane analysis.

The technique is designed to identify surface-breaking cracks in the bore using a standard eddy current probe, operating at 2 MHz. The technician has a number of issues to deal with in applying the technique: maintaining good probe contact; covering the entire surface of the cylinder; and finally, accurately mapping the faults to enable them to be blended out.

In this instance false calls will result in the unnecessary blending of primary structure that could have a significant effect on the airworthiness of the aircraft. The technique and equipment have been optimised to detect faults such that any fault detected must be reported. Additionally, post blending, the technician is required to undertake a final scan of the surface to prove the extremities of the fault have been removed. It is at this juncture that the technician is under the closest scrutiny as there is a high degree of probability that a new fault may reappear in the same area as the residual stress distribution alters, promoting the further development of microstructure cracking.

Nimrod 2000 NDT Programme. The Nimrod 2000 programme is the RAFs replacement maritime patrol aircraft. The project involves the refurbishment of existing airframes, with large proportions of the structure being retained. The retained structure is being subjected to a 100% NDT inspection programme to determine the extent of any corrosion, or any faults that may be present.

The objective of the project is to produce a structure that will effectively have been extended in life by 16 000 flying hours, or 25 years. Additionally, the aircraft is guaranteed to be corrosion-free for the first 18 years service. In total 21 aircraft will be refurbished over the next 10 years. Figure 1 shows the scale of the project.

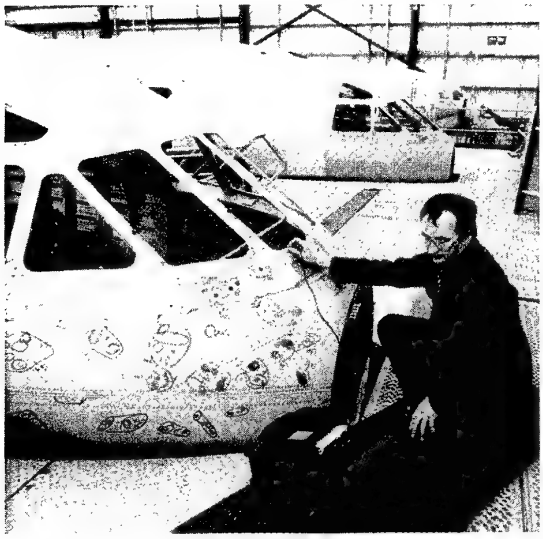


Figure 1

In order to undertake the NDT structural survey, a number of large-area scanning systems have been utilised. The first aircraft through the programme has been used as a development tool to establish which methods are best suited for the task. Consequently, all the faults that have been found have been validated by a secondary inspection method.

The fuselage skin was C-scan mapped using low frequency eddy currents, and the lap joints inspected using double-pass light diffraction techniques. In both instances the faults that were detected were validated using radiography. Additionally, the Redux bonded stringers are being examined using ultrasonic resonance techniques and C-Scan mapping the areas covered. Figure 2 shows the area scanning system used.

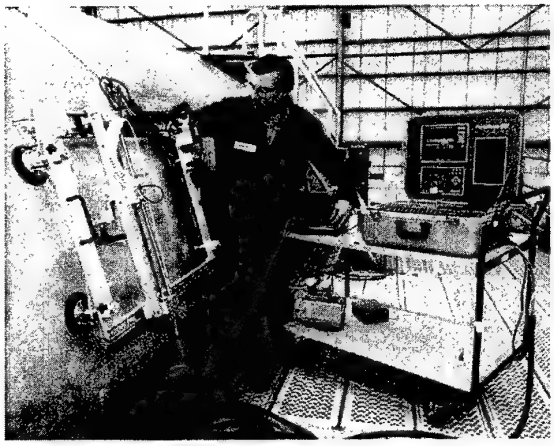


Figure 2

The refurbishment contract is with BAe, with NDT Squadron personnel acting as technical consultants to both the RAF project managers and BAe on the required inspection standards. Periodic audits are carried out to independently validate the areas inspected and, to date, the results of the audits have shown close correlation.

The implementation of large area scanning systems has proven to be invaluable in assessing the best methodology. C-Scan mapping provides an essential tool in imaging not only defect areas, but it also provides an audit trail of good structure. Therefore, we believe that any aircraft life-extension programme must fully utilise the technology that is available.

Avco Lycoming I/O 360. The Avco Lycoming I/O 360 engine (Figure 3) is used to power the Bulldog basic flying training aircraft.

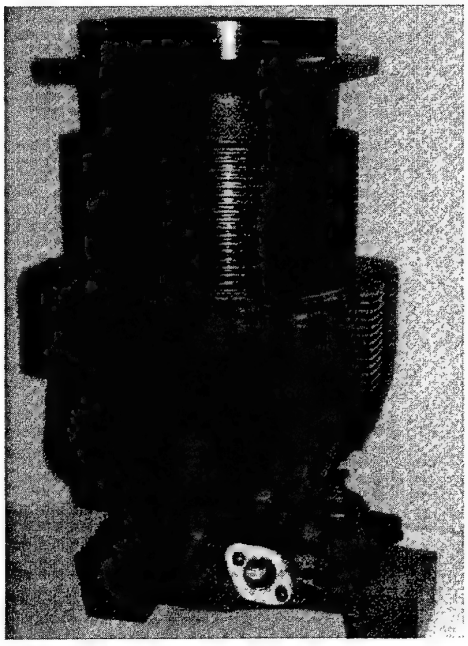


Figure 3

The Squadron was tasked to develop a technique to inspect the cylinder casting, following the catastrophic failure of a cylinder in flight. The geometry of the casting is such that a special-to-type probe was required. The probe developed was a twin crystal compressional probe with a beam angle of 12° in steel. In order to gain good contact with the inspection surface, the geometry of the body of the probe had to be manufactured to match the outside diameter of the cylinders' cooling fins. Additionally, there was minimal damping in the delay line of the probe as the probe had to fit into a 5mm high recess in the body of the casting/cylinder interface. The signal-to-noise ratio is relatively low, with a number of shear wave mode conversions taking place inside the casting, thus resulting in a large number of unwanted signals. We were, however, able to establish that the probe could clearly detect a 5mm deep radial fault, on calibration blocks with EDM slots cut into them. This was then accepted as the fault defect standard.

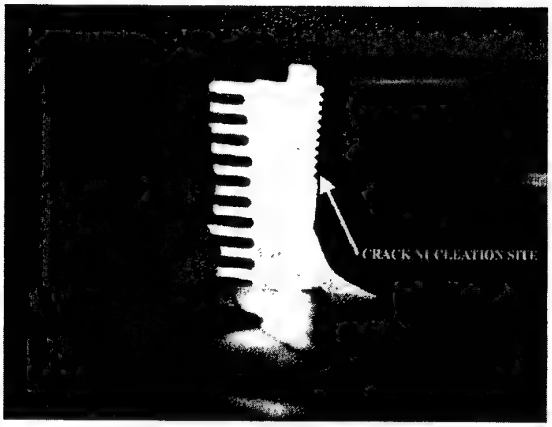


Figure 4

Figure 4 shows a cross sectional view of the cylinder and the probable depth location of crack nucleation. Prior to the technique being issued little information was available on the probable circumferential position of the crack nucleation site. However, one aspect of the technique which was an absolute certainty was that if a fault propagates radially to the outer circumference, the signal-to-noise ratio was increased by a factor of approximately 5, and the fault indication reached full screen height. Hence, should there be a gross crack present it would be positively identified with a very high degree of certainty. Unfortunately, anomalous signals could appear at particular points on the circumference as a result of mode conversions reflecting from the internal surfaces. Therefore, there was also a high probability of false calls being made. Notwithstanding both issues, the technique was able to certify engine cylinders that were good.

The inspection phase identified 25 cylinders as having fault indications from a total of 288 inspected. Initially 4 cylinders were sectioned with 2 being false calls and the other 2 having faults in them. DERA personnel produced C-Scan maps of the faults that are shown at figures 5 and 6.

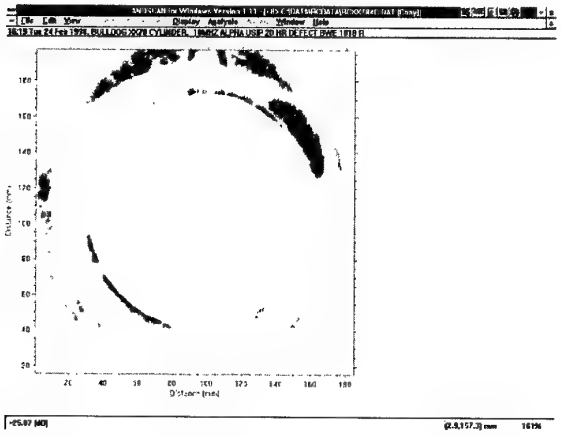


Figure 5

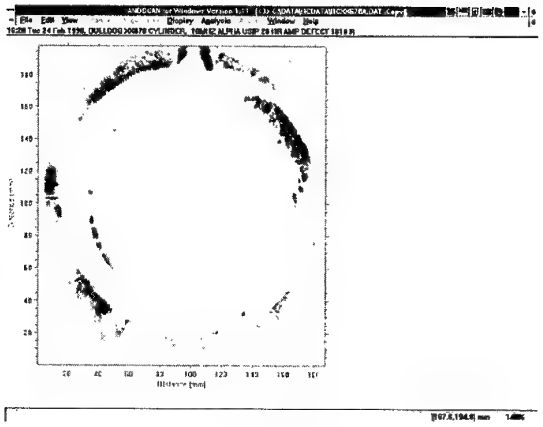


Figure 6

The reported fault size in both instances was only 1.5 cm, whereas the C-Scan images show that the actual fault size was approximately 6.0 cm in the circumferential direction (dark areas covering the 1-3 o'clock positions). Hence, the ultrasonic beam is not able to focus on the crack nucleation site. Further development work is now underway to determine the radial depth of the fault at which the signal-to-noise ratio is high enough to discriminate faults. Also, we are investigating if the beam angle can be increased without complete mode conversion, thereby focusing the probe more directly onto the probable crack initiation site and reducing the defect standard.

In this case study, the operational requirement to clear the aircraft for flight meant that we were unable to fully establish all the fundamental parameters of the technique. This resulted in a technique that may have an associated high false call rate. However, the technique was able to determine if cylinders were good, within the limitations of the inspection. Work has now commenced to evaluate the condition of the remaining 21 cylinders and produce a full POD evaluation of the technique. What is significant in this instance is that the study was initiated after a complete round of inspections had been undertaken. Hence, the technicians were not influenced by the dilemma of a reliability study and all the raw data was and remains uncorrupted.

The situation has provided NDT Squadron with a unique opportunity to undertake a POD analysis of a technique that is probably one of the most challenging we have undertaken in recent years. Results from the study will be used as a management aid to quantify the various effects each of the key factors had on the success of the technique. It is hoped that we will be able to improve the overall success of the technique by targeting the most significant factor in the equation.

6 CONCLUSIONS

In order to have confidence in the reliability and repeatability of an NDT inspection, the 3 key dependant

factors must be addressed. Utilising the best practice technique, within the economics of scale, and having highly motivated, well trained NDT technicians supported by a robust Squadron infrastructure optimises the reliability of any NDT technique.

All techniques require a degree of validation, whether that be a complete POD analysis or by manufacturing calibration and training standards specifically for the technique. By careful selection of the fault/defect standard, the deviation in reliability caused by variances in equipment are eliminated, leaving the main POD concerns centred on 'human factors'.

The utilisation of new technology must be carefully evaluated to ensure that in-service criteria are fully satisfied. Additionally, the latest generation of area-scanning equipment has simplified the interpretation skills of technicians and in some cases improved fault detection capabilities.

AIRFRAME INSPECTION RELIABILITY USING FIELD INSPECTION DATA

J.H. Heida and F.P. Grooteman
National Aerospace Laboratory NLR
P.O. Box 90502, 1006 BM Amsterdam
The Netherlands

1. SUMMARY

The possibilities within the Royal Netherlands Air Force (RNLAF) maintenance system to establish reliability data relevant for the in-service nondestructive inspection of F-16 airframe structure are described. The principal inspection techniques herewith are manual and automatic eddy current inspection for the detection of fatigue cracking. Use is made of field inspection data registered in the Core Automated Maintenance System (CAMS) for specific airframe inspection points within the F-16 Aircraft Structural Integrity Program (ASIP). The available data include the registration of the number of cracks and the length of the largest crack found during the phased inspections. Further, use is made of crack growth data obtained from the aircraft manufacturer. An evaluation of the field inspection data and the crack growth data allows the estimation of the sensitivity and reliability of inspection for the structural details concerned. The results of this evaluation can be used to revise the current values of the inspection intervals for the ASIP inspection points.

2. INTRODUCTION

Nondestructive inspection (NDI) is an integral part of aircraft maintenance. It is important to select the appropriate NDI techniques and to select the inspection times in terms of initial inspection (inspection threshold) and inspection interval, especially because of their impact on the balance between flight safety and maintenance costs. A too conservative maintenance approach could include unnecessarily frequent inspections resulting in high maintenance costs without an additional increase in flight safety. On the other hand, insufficient maintenance (inspection) could directly lead to an unacceptably low level of flight safety.

The selection of appropriate NDI techniques and the inspection frequency are related to each other because aircraft such as the F-16 have been designed in accordance with the Damage Tolerance (DT) design philosophy (Ref. 1). Damage Tolerance can be defined as "the ability of aircraft structure to sustain anticipated loads (e.g. limit load) in the presence of fatigue, corrosion or accidental damage until such damage is detected through inspections (or malfunctions) and repaired". In the DT design philosophy it is assumed that flaws already exist in the structure as manufactured, and that the structure may be inspectable or non-inspectable in service. Non-inspectable structures must be designed in such a way that the initial damage will not propagate to a critical size (causing failure) during the design service life. For inspectable structures the initial damage must grow slowly and not reach a critical size in some predetermined inspection interval.

The DT approach for inspectable structures is illustrated in figure 1. It is conservatively assumed that all specimens of a specified configuration contain an initial flaw (flaw size a_i) that propagates at a known rate. The assumed initial flaw size is small and generally not detectable with current inspection techniques. After a certain propagation time in service the flaw becomes reliably detectable (flaw size a_d) with a certain NDI technique. Finally, the critical flaw size a_c is assumed to be known from fracture toughness data; a_c is usually defined as the flaw size for which the structure can just sustain limit load.

The initial inspection time (I_1) and inspection interval (ΔI) are subsequently determined:

- I_1 is the flaw propagation period from a_i to a_c (this period is also called the "safety limit" SL) divided by a safety factor. This factor is usually taken as 2 which gives: $I_1 = ½ . SL$.
- ΔI is the flaw propagation period from a_d to a_c (period $\Delta = I_c - I_d$) divided by a safety factor. This factor is usually taken as 2 which gives: $\Delta I = ½ . \Delta$.

The relation between the appropriate NDI technique and the inspection frequency can now be understood. Visual inspection or low level NDI inspection are low cost inspection methods but have a relatively large detectable flaw size a_d and consequently a short inspection interval ΔI. On the other hand, a more advanced NDI technique is more costly in application but will have a smaller a_d and, consequently, will have a larger inspection interval. The aircraft operator has then the choice between frequent inspections with relatively high a_d inspection techniques or less frequent inspections with relatively small a_d inspection techniques, both yielding a same level of cumulative reliability of inspection.

In this paper first some general aspects of NDI reliability will be discussed. Then, the possibilities within the Royal Netherlands Air Force (RNLAF) maintenance system to establish reliability data, especially a_d values, relevant for the in-service nondestructive inspection of F-16 airframe structure will be described. Use will be made of field inspection data registered in the Core Automated Maintenance System (CAMS) for specific airframe inspection points.

3. RELIABILITY OF NONDESTRUCTIVE INSPECTION

The reliability of NDI is generally associated with the ability of an inspector to detect flaws in the parts inspected. The probability of detection (POD) for the flawed parts is then usually taken as measure of the inspection performance. The true POD for a particular flaw size, however, can only be obtained by means of an infinite number of inspections. In practice, a limited number of inspections will only yield an estimated POD. To provide a measure of confidence in the estimated POD, it is usual to incorporate confidence limits (CL) resulting in lower-bound values of the POD. An often quoted value for the reliably detectable flaw size is the 90/95

Paper presented at the RTO AVT Workshop on "Airframe Inspection Reliability under Field/Depot Conditions", held in Brussels, Belgium, 13-14 May 1998, and published in RTO MP-10.

POD/CL value i.e. the flaw size for which we have 95 % confidence that the true POD is 90 % or more.

In practice, however, the majority of the specimens inspected are without flaws, yielding the possibility of obtaining a spurious indication of a non-existing flaw. Hence, the result of an inspection can be described by means of a quadrinomial distribution, with successful and unsuccessful inspections of both flawed and unflawed specimens (Fig. 2). In analogy with the POD for the flawed specimens a probability of recognition (POR) can be defined for the unflawed specimens. Also for the POR, lower confidence limits can be calculated with statistical methods. Often, the counterpart of POR viz. the false calls probability (FCP) is used as inspection characteristic for the unflawed parts. Both POD and POR (or FCP) are essential inspection characteristics with their relative importance depending on considerations of safety and economy (Ref. 2). An attractive way to visualize the inspection performance is a diagram in which the POD and POR (or FCP) values are plotted against each other as the detection threshold is varied, yielding a so-called "relative operating characteristic" or ROC curve (Ref. 3). Such a diagram can be useful, for example for the comparison of different inspection techniques and for the performance ranking of individual inspectors.

In this paper we will focus on the POD for flawed parts because this is the most important inspection characteristic from a safety point of view.

4. NDI RELIABILITY DEMONSTRATION

Independent of the definition of the reliably detectable flaw size a_d, e.g. the 90/50 or 90/95 POD/CL value, one has to determine the POD curve of the relevant NDI techniques for a specific inspection configuration (specimen configuration), see figure 3. For this purpose a so-called NDI reliability demonstration program can be performed.
The design of such a program has been well addressed in an AGARD SMP Lecture Series (Ref. 4). This document describes testing and evaluation procedures for assessing the capability of an NDI system in terms of POD and confidence limits. NDI systems are herewith classified into two categories depending on the outcome of an inspection: NDI systems which produce only qualitative information as to the presence or absence of a flaw ("hit/miss" data) and NDI systems which record a signal response [â] that is correlated with the actual size [a] of the indicated flaw ("â vs. a" data). For both NDI systems, reference 4 gives recommendations for modelling the POD and for calculating lower confidence bounds.
The design of a reliability demonstration program has also been addressed in an FAA supported project at the Aging Aircraft NDI Development and Demonstration Center in Albuquerque (Ref. 5). The three-volume document presents a generic protocol for the conducting of inspection reliability experiments, it further presents a specific protocol for an eddy current inspection reliability experiment, and it gives the results of an actually performed reliability experiment at different airline inspection facilities for the manual high-frequency eddy current inspection of aircraft lap splice joints. Topics addressed include the presentation of POD curves, the treatment of false calls and the presentation of ROC curves.
Further, the NRC Institute for Aerospace Research (IAR) in Canada has performed extensive NDI reliability studies and experiments. For example reference 6 gives the results of an AGARD round-robin NDI demonstration program in which six laboratories in four NATO countries participated. In this program several NDI procedures were evaluated for the inspection of bolt holes of service-expired compressor disks and spacers from the J85-CAN40 engine.
A reference book of available quantitative NDI data has been compiled by the NTIAC in Austin (Ref. 7). This reference book gives guidelines for demonstration of specific NDI process capabilities and it provides more than 400 POD curves for various NDI techniques applied for various inspection configurations.

A well performed NDI reliability demonstration program can yield the necessary reliability data, for example a_d values, for a certain inspection configuration. However, such programs also have their limitations. Besides representativity of inspection configuration and the influence of human factors, the main limitations of performing an NDI reliability demonstration program are the time and costs involved. Especially the number of test specimens necessary for the "reliable" determination of POD and ROC curves is very large. For example, reference 4 recommends that the specimen set should contain at least 60 flawed sites if the NDI system provides only "hit/miss" results and at least 40 flawed sites if the NDI system provides a quantitative response, "â vs. a" data. Furthermore, to enable the estimation of the false call rate, reference 4 recommends that the specimen set should contain at least three times as many unflawed inspection sites as flawed sites.

These limitations are the reason that NDI reliability demonstration programs are infrequently performed and then for applications with only one or with a limited number of inspection configurations. When a large number of different configurations is involved, as for NDI of airframe structure, it is impractical to conduct these extensive programs for each different structural detail. Different approaches can then be distinguished:

- Conduct a limited number of NDI reliability demonstration programs on selected structural details and extrapolate the results of these programs to comparable structural details.
- Make a conservative use of available data from the literature, for example of relevant POD curves from the NDE capabilities data book (Ref. 7).
- Make use of field inspection data e.g. the NDI results of in-service fleet inspections.

The last approach is an attractive option because of the acquisition of relevant results and because of the relatively low costs involved. Therefore, the "field data use" approach will be further discussed for the in-service NDI of F-16 airframe structure within the Royal Netherlands Air Force.

5. RNLAF IN-SERVICE NDI OF F-16 AIRFRAME STRUCTURE

The general NDI procedures for the in-service inspection of the F-16 airframe structure are described in reference 8. A RNLAF supplement on this reference lists the specific inspection control points within the F-16 Aircraft Structural Integrity Program (ASIP). In this paper the attention will be focused on the ASIP control points because of the crack growth information available (e.g. crack growth curves, critical flaw sizes) and because of the use of a comprehensive registration system for the ASIP field inspection data i.e. the Core Automated Maintenance System (CAMS). When cracks

are detected during the inspection of an ASIP point, then the number of cracks and the length of the largest crack found (amongst other general data) are registered in the CAMS system.

The values for initial inspection time (I_1) and inspection interval (ΔI) for each ASIP point are listed in the Fleet Structural Maintenance Plan (FSMP) for RNLAF F-16 aircraft. The I_1 and ΔI values have been derived using the Damage Tolerance approach explained in chapter 3 (I_1 = ½ . SL and ΔI = ½ . Δ), using fatigue crack growth curves relevant for RNLAF usage (determined with a load spectrum based on the actual RNLAF base usage) and using reliably detectable flaw size a_d values based on assumed in-service NDI capability.
The following a_d values are currently used for the primary NDI procedures of ASIP points (all flaw sizes relate to surface crack lengths):

- Manual eddy current inspection: a_d = 0.10, 0.20 or 0.25 inch, depending on the inspection location.
- Automatic eddy current inspection (rotating probe) of bolt holes: a_d = 0.075 inch.
- Magnetic particle inspection: a_d = 0.10 inch.

The majority of the ASIP primary inspections include manual and automatic eddy current inspection. Magnetic particle inspection is only applied for a small number of ASIP points (e.g. the canopy hook support fitting). Penetrant inspection is only used as a back-up NDI procedure. Ultrasonic inspection is applied for a number of inspection points (e.g. the shock strut piston radius of the nose landing gear) but these inspections are RNLAF specific.

Up to now, the a_d values for the ASIP inspection points have been based on assumed in-service NDI capability. These a_d values seem conservative when compared with values from the literature. This means that the inspection intervals for these points may be unnecessarily conservative (large). Therefore, it is worthwhile to evaluate the available field inspection data in CAMS and to assess realistic a_d data for the ASIP inspection points. This information can then possibly be used to revise the current values of the ASIP inspection intervals.

6. POD ASSESSMENT USING FIELD INSPECTION DATA

The CAMS system registers the number of cracks and the length of the largest crack found during the inspection of an ASIP point. The NDI signal responses [â] are not recorded, so the NDI data base is of the "hit" type. Information of the sizes of undetected cracks ("miss"), however, is necessary for the construction of a POD curve (analysis of "hit/miss" data). But, when crack growth data are available, for each crack detected the previously missed crack sizes (during previous inspections) can be estimated (Refs. 9, 10). When crack growth data are not available, the data base will only contain crack detection data. These data can then be used in a limited approach to estimate a_d values by plotting a Cumulative Distribution Function of the crack sizes detected.

Crack growth data available

For most ASIP inspection points crack growth data are available. These data include realistic crack growth curves and values for the critical crack size a_c. The crack growth curves can be used to estimate the previously missed crack sizes for each crack detected during an inspection. This procedure is illustrated in figure 4. When this procedure is applied for the inspection of an ASIP point for all aircraft in service, this will result in an NDI data base of the "hit/miss" type for that particular ASIP point. When sufficient data are available (see chapter 4) a POD curve can be constructed. In the literature different models of a POD curve for the analysis of "hit/miss" data have been suggested. The most appropriate POD models have been evaluated by the NRC/IAR using the inspection results of actual aircraft engine disks containing service-induced cracks (Ref. 11). It was concluded that the log-normal regression function provides the most realistic POD results. This function was also recommended in an AGARD SMP Lecture Series (Ref. 4).
The log-normal model to relate the POD with crack size [a] can be formulated as follows (after Ref. 4):

$$POD(a) = 1 - Q(z)\ ;\ z = (\ln(a) - \mu)/\sigma \qquad (1)$$

where Q(z) is the standard normal survivor function, z is the standard normal variate, and μ and σ are the location (mean) and scale (standard deviation) parameters.
The two parameters (μ, σ) must be determined with a parameter estimation procedure. Also here, different methods have been mentioned in the literature, such as the Maximum Likelihood Estimators (MLE) method and the Range Interval Method (RIM). These methods have been evaluated in reference 11; it was concluded that the MLE method is the preferred method. For example, the MLE method does not require any information other than the actual "hit/miss" data.
An example of the construction of a POD curve from "hit/miss" data following the aforementioned method (log-normal POD function, MLE parameter estimation procedure) is given in figure 3 (from Ref. 6). This figure gives the mean POD curve (50 % confidence) and the lower-bound POD curve with a 95 % confidence level. In this example the reliably detectable flaw size a_d has been defined as the 90/95 POD/CL value yielding a 2.6 mm crack length.

Crack growth data not available

For some inspection points crack growth data may not be available. In that case it is not possible anymore to estimate the previously missed crack sizes for each crack detected during an inspection. It is also not possible then to construct a POD curve from the available "hit" data. However, the crack detection data can still be used in a limited approach to obtain information about the detectable crack size by constructing a detection threshold histogram (Ref. 10). For this purpose, the available data are grouped in appropriate intervals of detected crack size, and a histogram is made of the frequency of detection versus crack size. The histogram can yield information such as the sensitivity of inspection (detection threshold) and the mean crack size detected.
A further approach is to assume a Probability Density Function (PDF) for the crack sizes detected and to calculate its integral i.e. the Cumulative Distribution Function (CDF). In analogy with the aforementioned POD(a_i) calculation (with both "hit" and "miss" data available) a log-normal PDF is assumed for the crack sizes detected ("hit" data):
where μ and σ are the mean and standard deviation of the log crack sizes detected.
Next, a Cumulative Distribution Function (CDF) can be constructed, indicating the probability that the detected crack

$$\text{PDF:}\quad f(a) = \frac{1}{a\sigma\sqrt{2\pi}}\, e^{-\frac{1}{2}\left(\frac{\ln(a)-\mu}{\sigma}\right)^2} \qquad (2)$$

size has a value less than or equal to [a]:

$$\text{CDF:}\quad F(a) = \int_{x=0}^{x=a} f(x)\,dx \qquad (3)$$

To illustrate the PDF/CDF approach the inspection data of figure 3 (from Ref. 6) have been reviewed. The data comprise 79 "hits", 206 "misses" and only 1 false call. The PDF and CDF for the "hit" data are shown in figure 5. The mean and standard deviation are 2.3 mm and 1.2 mm crack length, respectively. These parameters have been determined with the least squares estimates procedure (in fact, first the μ and σ of ln(a) have been calculated). The goodness-of-fit for the data is shown in figure 6. Figure 5 allows an estimation of the detection threshold (about 0.5 mm) and of the crack length [a] for which there is a 90 % probability that the detected cracks have a length less than or equal to [a] (a = 3.8 mm). The reliably detectable crack length a_d can not be extracted from the CDF.

In figure 7 both the CDF for the "hit" data and the mean POD curve (confidence level 50 %) for the "hit/miss" data from figure 3 have been drawn. A 90 % probability criterion yields the crack lengths 3.8 mm and 2.4 mm for the CDF and POD curve, respectively. These values can not be compared directly: 3.8 mm is the crack length for which there is a 90 % probability that the detected cracks have a length less than or equal to this 3.8 mm, while 2.4 mm is the crack length for which there is a 90 % probability of detection (the 90/95 POD/CL value is a 2.6 mm crack length).
In general, the 90 % probability flaw size calculated from a CDF will be larger than the flaw size with a 90 % probability of detection. It can be concluded that the CDF can not give an exact value of the reliably detectable flaw size a_d, but it can give a conservative estimate of this a_d.

7. RNLAF F-16 FIELD INSPECTION DATA

The CAMS system registers the field inspection data of about 65 ASIP points in F-16 aircraft. At the moment there is an extensive CAMS data base but the amount of data crack detection is still limited because:

- Some ASIP points have large inspection intervals (e.g. exceeding 1000 flight hours) and hence acquire few inspection data.
- For a large number of ASIP points (almost) no cracks are detected.
- For some ASIP points the available crack detection data are the result of a first inspection, so that information of previously missed crack sizes can not be extracted.
- For some ASIP points the CAMS data base has not been kept up with completely (e.g. discipline of data filling-out).

The result is that at the moment for only a few ASIP points a sufficient number of crack detection data is available from which a relevant "hit/miss" data base can be deducted. As an example, ASIP control point 3005 will be taken to show the intrinsic possibilities of further analysis of field inspection data.

ASIP 3005 deals with the inspection of the tab radii in the F-16 16B5120 center fuselage longeron, see figure 8. The longeron is a tee-extrusion machined from 2024-T62 aluminium, and functions to distribute flight loads from the fuselage upper skin to the center fuselage structure. High positive g-loads may cause fatigue cracking in the tab radii of the longeron. NDI involves a manual eddy current inspection technique using a standard eddy current phase-analysis instrument and a 50-200 kHz shielded pencil-probe (Ref. 8). The current value for the reliably detectable crack size a_d has been set at a through-crack (0.090 inch plate thickness) with a length of 0.10 inch.
The crack growth curve for the ASIP 3005 control point is shown in figure 9. It is in fact a durability crack growth curve with an initial flaw size of 0.007 x 0.007 inch and a functional impairment crack size of 0.187 inch. Durability is not a safety life concept but an economic life concept; the durability life represents the life for which flaws will not grow to an extent that requires extensive repair before one design service life. ASIP 3005 is treated as a durability item (and not as a damage tolerance item) because the 16B5120 longeron is believed not to be a safety of flight structure; the predicted durability life is 4320 flight hours (Ref. 12).
The current inspection interval is 200 flight hours; it is in fact not based on the crack growth data of figure 9 but on a former durability analysis of the aircraft manufacturer using a different crack growth curve. That analysis resulted in a relatively short interval (less than 100 flight hours) which was rounded up to a phase inspection interval of 200 flight hours, however, because of the longeron not being a safety of flight structure.

The available CAMS field inspection data of ASIP 3005 are given in table 1. This table lists for 27 aircraft the actual crack lengths detected and an estimation of the crack lengths missed during the previous inspections (between brackets). For this crack length estimation the crack growth curve in figure 9 was used. It is possible that in practice some cracks have been missed and which are hence not included in table 1. This will however only influence the size of the NDI data base and not significantly the shape of the POD curve (and a_d assessment). In total, the inspection results yield 28 "hit" data points and 36 "miss" data points (in total 64 "hit/miss" data points). These data points have been used to draw a CDF and a mean POD curve, see figure 10. The two curves correlate remarkably well and show that the sensitivity of inspection (detection threshold) is about 0.02 inch (0.5 mm). Further, a 90 % probability criterion yields the crack lengths of 0.093 inch (2.4 mm) and 0.108 inch (2.7 mm) for the POD and CDF curve, respectively. Without defining a specific confidence level on the POD to determine the reliably detectable crack size a_d, the POD curve in figure 10 indicates that the a_d value lies in the range of 0.10 inch. This value is equal to the currently used value of a_d for ASIP 3005 and for other comparable ASIP points inspected with the manual eddy current technique. Finally, it is emphasized again that the CDF can not give an exact value but only a conservative estimate of a_d.

8. CONCLUDING REMARKS

In this paper the possibilities within the RNLAF maintenance system to establish reliability data relevant for the in-service nondestructive inspection of F-16 airframe structure have been

described. It has been shown that an evaluation of the CAMS field inspection data and crack growth data allows an estimation of the sensitivity and reliability of inspection for the structural details concerned. The results of such an evaluation can be used to revise the current values of the ASIP inspection intervals.

For the ASIP 3005 inspection point it has been shown that the reliably detectable flaw size lies in the same range as the currently used value of a_d (0.10 inch). So, in this particular case no revision of the currently used inspection interval is proposed. It is nevertheless a remarkable outcome because it has often been suggested that the value of 0.10 inch is on the very conservative side for this inspection configuration. A quick survey of the field inspection data in table 1 does also suggest this. The lesson learned is thus that realistic values for a_d are often larger than generally assumed.

The ASIP 3005 evaluation has demonstrated that the CAMS field inspection data can, in principle, be used to determine more realistic a_d values and hence more realistic values of the ASIP inspection intervals. For most ASIP points, however, the a_d and ΔI evaluation can not yet be performed because of the limited amount of crack detection data in the CAMS data base, see chapter 7. Some possibilities to overcome this limitation are:

- Stringent maintenance of the CAMS data base.
- Combination of crack detection data for ASIP points with comparable inspection configuration such as location and inspection technique (for example for the carry-through bulkhead ASIP points).
- Combination of RNLAF crack detection data with comparable crack detection data of other Air Forces. Estimation of previously missed crack sizes can then be done using crack growth curves incorporating a Crack Severity Index (CSI) for differences in base usage (load spectrum).

For the last item it is recommended to perform this activity within the framework of a NATO RTO Working Group to be established.

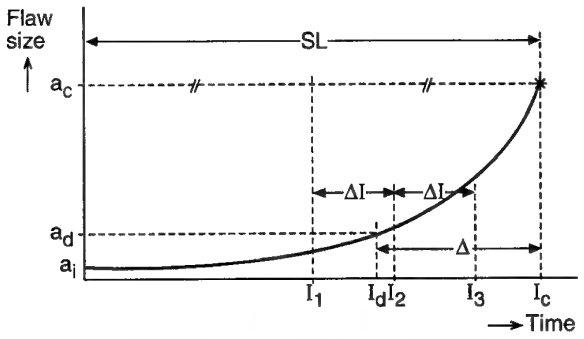
9. REFERENCES

1. Military Specification, "Airplane Damage Tolerance Requirements", MIL-A-83444 (USAF), 2 July 1974.
2. Heida, J.H., "Quadrinomial distribution for the characterization of NDI reliability", NLR report MP 84064 U, Proceedings of the 3rd European Conference on Nondestructive Testing, Florence, Italy, 15-18 October 1984.
3. Swets, J.A., "Assessment of NDT systems; Part I and II", Materials Evaluation, Vol. 41, No. 11, pp. 1294-1303, October 1983.
4. Petrin, C.L., Annis, C., and Vukelich, S.I., "A recommended methodology for quantifying NDE/NDI based on aircraft engine experience", AGARD Lecture Series 190, April 1993.
5. Spencer, F. et al, "Reliability assessment at airline inspection facilities; Volumes I, II and III", FAA report DOT/FAA/CT-92/12, 1993 (I and II) and 1995 (III).
6. Fahr, A. et al, "POD assessment of NDI procedures using a round robin test", AGARD report AGARD-R-809, January 1995.
7. Nondestructive Testing Information Analysis Center (NTIAC), "Nondestructive evaluation (NDE) capabilities data book", 3rd ed., NTIAC report DB-97-02, Austin, November 1997.
8. Lockheed Martin Corporation, "Nondestructive inspection; USAF/EPAF series F-16A and F-16B aircraft", Technical manual T.O. 1F-16A-36, Change 38, 3 March 1997.
9. Brewer, J.C., "Estimate of probability of crack detection from service difficulty report data", FAA report DOT/FAA/CT-94/90, September 1994.
10. Simpson, D.L., "Development of non-destructive inspection probability of detection curves using field data", NRC/IAR report LTR-ST-1285, Ottawa, August 1981.
11. Fahr, A., Forsyth, D.S., and Bullock, M., "A comparison of probability of detection (POD) data determined using different statistical methods", NRC/IAR report LTR-ST-1947, December 1993.
12. Lockheed, Fort Worth Company, "Durability and damage tolerance analyses for EPAF F-16A/B Block 10/15 mid-life update aircraft", report 16PR11666, Fort Worth, 27 August 1993.

Table 1 Available CAMS field inspection data of ASIP 3005; inspection of the tab radii in the F-16 16B5120 center fuselage longeron.
Listing of actual crack length [inch] detected and estimation of crack lengths missed during previous phased inspections (between brackets).

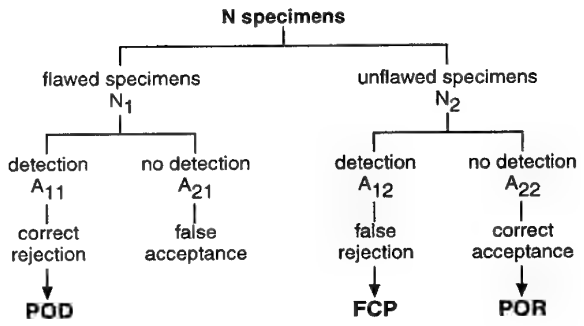
Aircraft CAMS Code	Phased Inspection Times (Flight Hours)								
	1200	1400	1600	1800	2000	2200	2400	2600	2800
A1869 RH	-	-	-	-	-	0.049			
A1870 LH	-	-	-	-	-	0.03			
A1870 RH	-	-	-	-	-	0.11			
A1871 RH	0.05								
A1873 RH	-	-	0.049						
A1874 RH	-	-	-	-	-	(0.029)	(0.035)	0.047	
A1875 RH	-	-	-	(0.025)	0.03				
A1876 RH	-	-	-	-	-	(0.038)	0.05		
A3199 RH	-	(0.019)	(0.021)	(0.025)	0.03				
A3202 RH	-	-	-	(0.025)	0.03				
A3203 RH	-	-	(0.021)	(0.025)	0.03				
A3204 RH	-	-	(0.019)	(0.021)	(0.025)	0.03			
A3208 RH	-	-	-	(0.025)	0.03				
A3209 RH	-	-	(0.021)	(0.025)	0.03				
A3616 RH	-	-	-	-	-	(0.030)	0.039		
A3620 RH	-	-	-	-	-	(0.026)	(0.031)	0.04	
A3623 RH	-	-	-	-	-	(0.019)	(0.021)	(0.025)	0.03
A3624 RH	-	-	-	-	(0.031)	0.04			
A3643 RH	-	-	-	(0.025)	0.03				
A3657 RH	-	-	-	-	-	(0.064)	0.15		
A4360 RH	-	-	(0.038)	0.05					
A4361 RH	-	-	(0.025)	0.03					
A4362 RH	-	-	(0.025)	0.03					
A5136 RH	-	(0.021)	(0.025)	0.03					
A5137 LH	-	-	(0.021)	(0.025)	0.03				
A8213 LH	-	-	(0.065)	0.157					
A8255 LH	-	-	-	-	(0.017)	0.019			
A8267 LH	-	-	-	-	-	-	(0.053)	(0.080)	0.236

Dash (-) means: no inspection data available



a_i = initial flaw size	SL = safety limit
a_d = reliably detectable flaw size	I_1 = initial inspection time (e.g. ½·SL)
a_c = critical flaw size	Δ = available inspection time
	ΔI = inspection interval (e.g. ½·Δ)

Fig. 1 Damage tolerance approach for inspectable structures.
Determination of the initial inspection time I_1 and the inspection interval ΔI



- Probability of detection : POD = A_{11}/N_1
- Probability of recognition : POR = A_{22}/N_2
- Probability of false calls : FCP = A_{12}/N_2

Fig. 2 The four possible outcomes of an inspection

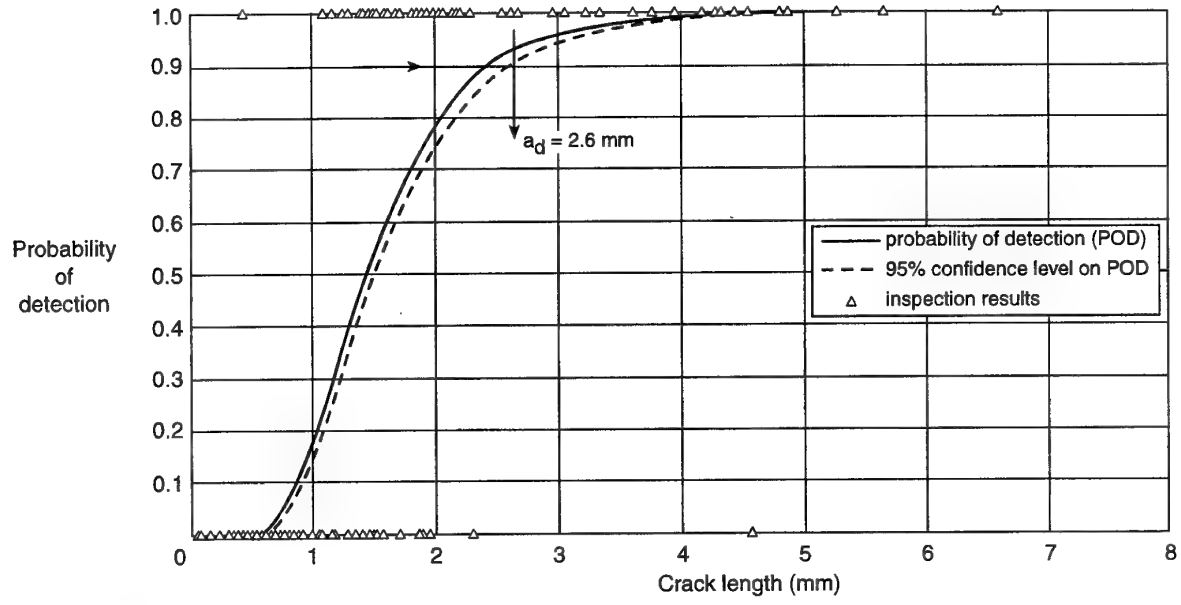


Fig. 3 Construction of a probability of detection (POD) curve, with its lower 95% confidence bound, from "hit/miss" data [Fig. 16 from Ref. 6].
Log-normal POD model with MLE parameter estimation procedure.
Reliably detectable flaw size a_d is 2.6 mm (here defined as the 90/95% POD/CL flaw size)

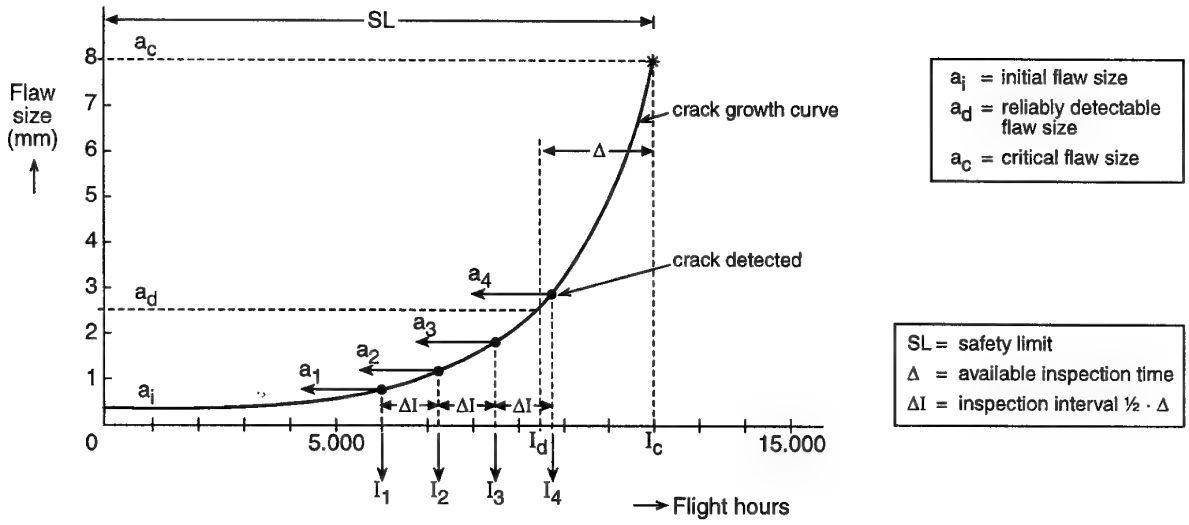


Fig. 4 Crack growth curve for a fictive ASIP control point, with the crack detected at the 4th inspection. Estimation of the crack sizes missed during the previous inspections I_1 (initial inspection), I_2 and I_3

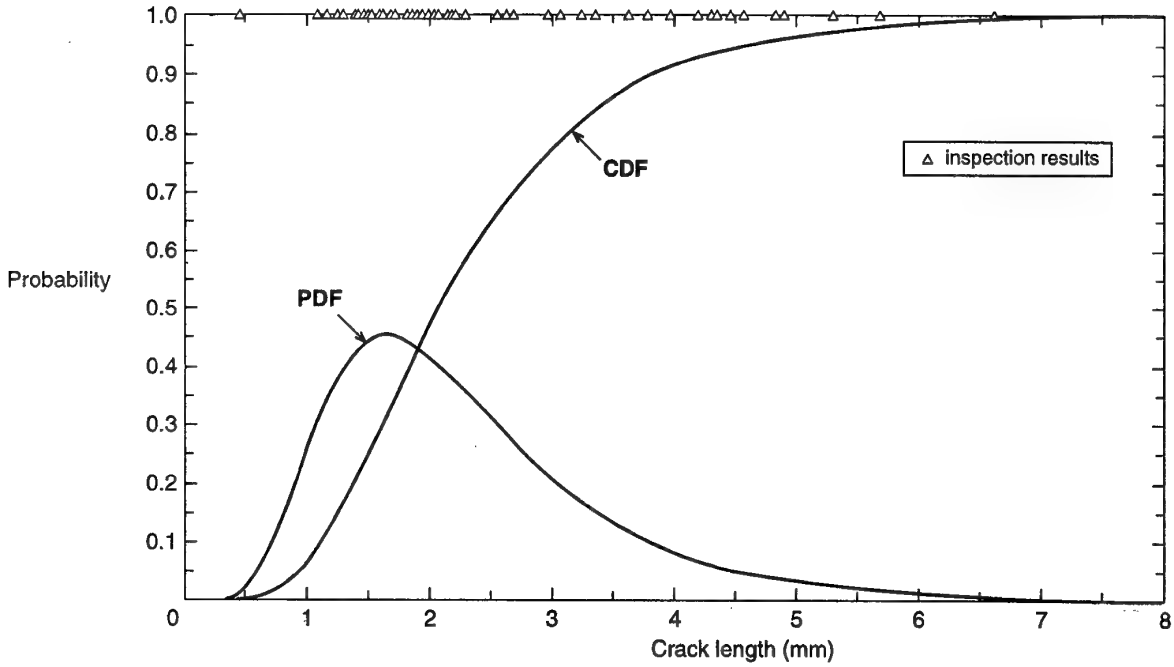


Fig. 5 Probability Density Function (PDF) and Cumulative Distribution Function (CDF) for the "hit" data (79 cracks detected) from figure 3

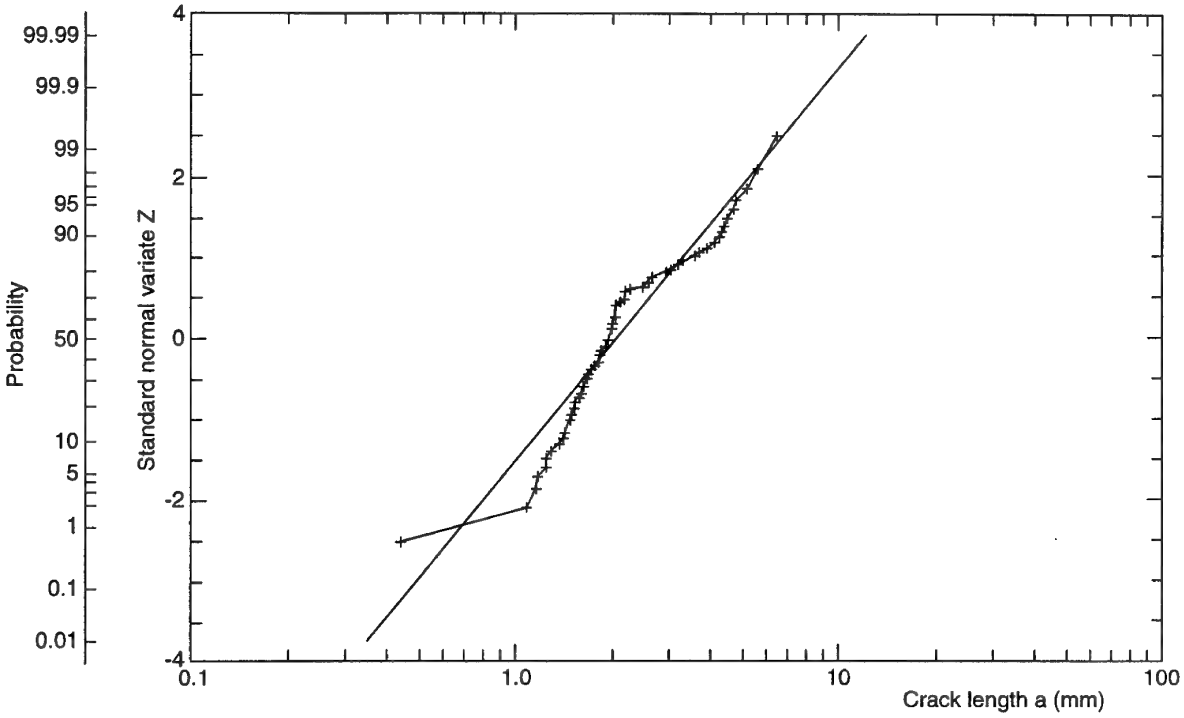


Fig. 6 Goodness-of-fit for the log-normal PDF estimation for the "hit" data from figure 3. Standard normal variate $z = (\ln(a)-\mu) / \sigma$ and its corresponding cumulative probability versus the crack length detected, plotted on log-normal probability paper

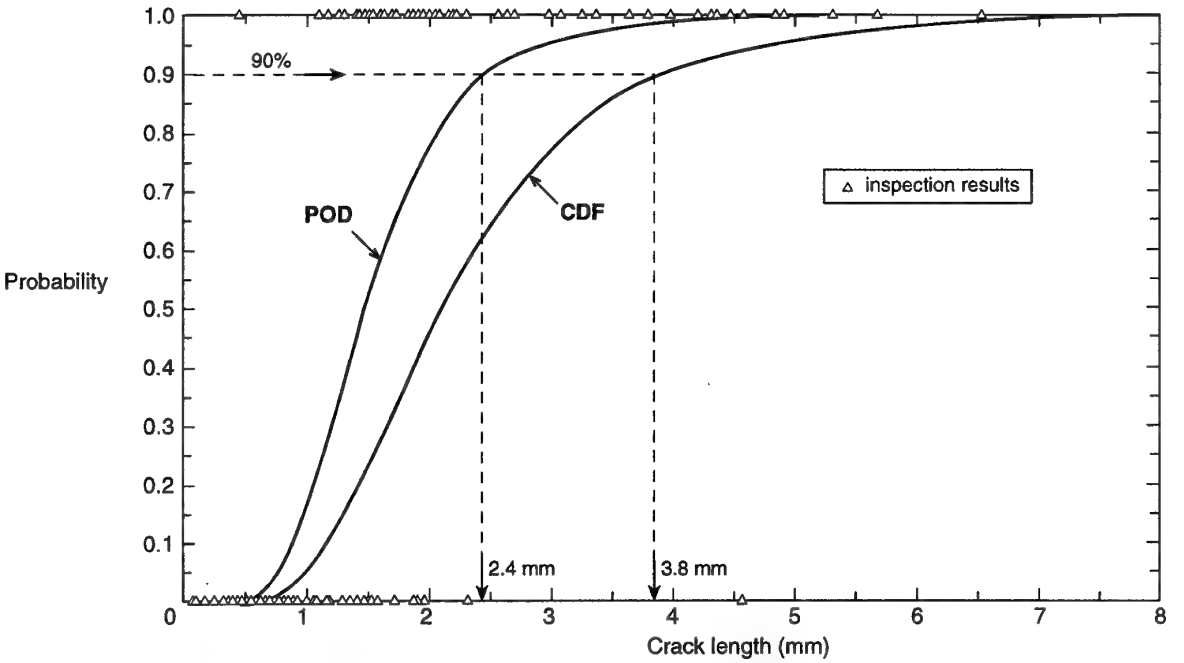


Fig. 7 Cumulative Distribution Function (CDF) for the "hit" data and Probability of Detection curve (POD, 50% confidence level) for the "hit/miss" data from figure 3

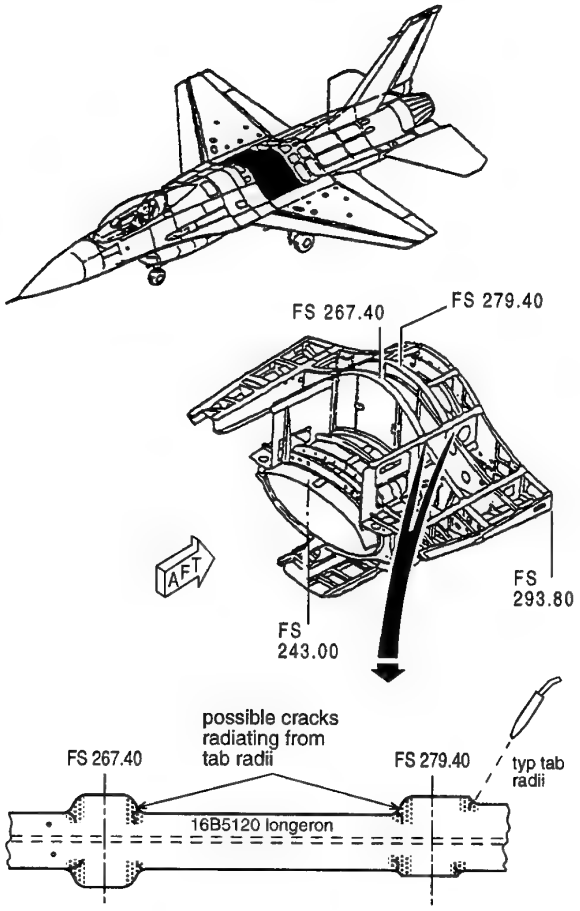


Fig. 8 F-16 ASIP 3005 inspection point. Manual eddy current inspection of the tab radii in the center fuselage longeron [Fig. 6-12 from Ref. 8]

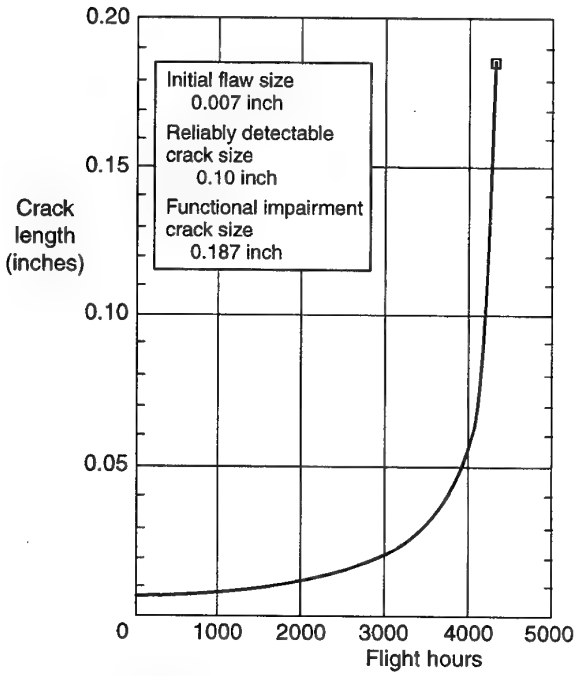


Fig. 9 Crack growth curve for the F-16 ASIP 3005 control point [Fig. 8.2.2-2 from Ref. 12]

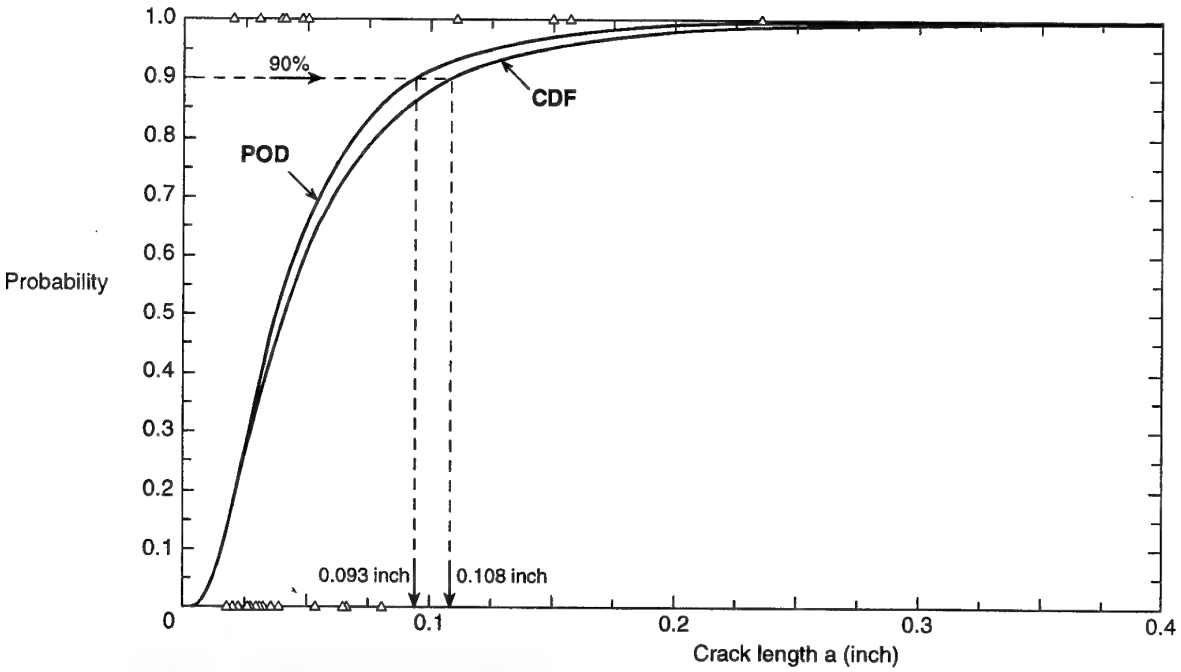


Fig. 10 CDF of 28 "hit" data points and mean POD curve of 64 "hit/miss" data points for the inspection of ASIP 3005, tab radii in the F-16 center fuselage longeron

PERFORMANCE EXPERIENCE AND RELIABILITY OF RETIREMENT FOR CAUSE (RFC) INSPECTION SYSTEMS

Sara Keller
OC-ALC/LPPEE
3001 Staff Drive 2B93
Tinker AFB, OK 73145-3034, USA

Carlos Pairazaman
VERIDIAN/VEDA Operations
Bldg 3001/IAMU Post Q95
Tinker AFB, OK 73145, USA

AL Berens
University of Dayton Research Institute
300 College Park
Dayton, OH 45469-0127, USA

Charles F. Buynak
US Air Force
Wright Laboratory
2230 Tenth St. STE 1
Wright-Patterson AFB, OH 45433-7817, USA

Robert Garcia
OC-ALC/LASAC
3001 Staff Drive
Tinker AFB, OK 73145-3071, USA

1. SUMMARY

The US Air Force Inspection Based Life Management of engine components requires an extensive Nondestructive Inspection (NDI) system Reliability Assessment. When this inspection technology is implemented in a production mode of operation, trade-off between better Probability of Detection - POD (lower thresholds) and throughput requirements become a way of life. Compromises between inspection requirements and "real Life" take place. The US Air Force experience developing, testing, and implementing Automated inspection systems, NDE technology, and Reliability testing are discussed.

2. INTRODUCTION

In the Aircraft industry, there are two predominant philosophies used to manage the life of engine components: The Conventional fatigue life design and the Damage Tolerance Approach (DTA). The first one, designs the engine component to the Low Cycle Fatigue (LCF) minus 3 Sigma limit and is based on the premise that all materials are free of initial defects. This philosophy makes no special allowance for material or in-process manufacturing anomalies or defects. Consequently, the NDE methods used for this approach are generally used as process control tools and require no detailed knowledge of detectable flaw sizes, or of the Probability of Detection (POD) for a particular flaw size. DTA assumes that damage, in the form of a flaw of minimum detectable size, is present within the component at all critical locations.

DTA concepts are used in the US Air Force Engine Structural Integrity Program (ENSIP) and the Retirement For Cause (RFC) program. These maintenance philosophies ensure that this fracture critical flaw will not grow to critical size in two inspection intervals, forcing inspections in manufacturing and at intervals of 1/2 average propagation life. Consequently, these philosophies demand a rigorous approach to assessing NDE capability. Quantitative, and statistically based NDE capabilities results are needed and have to be generated for the NDE techniques on the materials for which these will be used. For this reason, NDE detectable flaw sizes are defined recognizing these variables. Flaw size detection capabilities are expressed in terms of the POD, which is reported with two numbers, the calculated POD, and the Statistical Confidence Level associated with the calculation of that Probability.

3. THE ENSIP PHILOSOPHY

ENSIP is an organized and disciplined approach to the structural design, analysis, development, production, and Life Management of Gas Turbine Engines with the goal of ensuring: Engine structural safety, increase service readiness, and reduce life cycle cost. The roots of ENSIP lie in structural deficiencies and lost aircraft due to engine failures from all US engine manufacturers:

Paper presented at the RTO AVT Workshop on "Airframe Inspection Reliability under Field/Depot Conditions", held in Brussels, Belgium, 13-14 May 1998, and published in RTO MP-10.

- o 1946 Early turbine engines had 25 hours of operating life.
- o 1952 Turbine engines up to 160 hours of life.
- o 1960's Continued inadequate life capabilities.
- o 1969 First application of ENSIP concept without DTA.
- o 1973-74 Aircraft lost due to engine (without ENSIP concept) structural deficiencies. ENSIP concept formally introduced (still no DTA).
- o 1975-76 Aircraft (without ENSIP concepts) lost due to engine structural deficiencies.
- o 1978 ENSIP concept introduced with DTA.
- o 1979 First "ENSIP Assessment" on P&W F100
- o 1980 Similar assessment on GE TF34 and F101
- o 1982 ENSIP introduced as MIL Prime Specification
- o 1983 ENSIP requirements defined in MIL-SPEC-1783. New engine design requires ENSIP concepts. ENSIP concepts are also being utilized to manage fielded engines.

4. THE RFC PHILOSOPHY

The RFC philosophy extends the gas turbine engine components original design service life. The usefulness of engine components is based on predicted LCF crack growth behavior. The USAF uses a very conservative approach of retiring components after they reach a given number of operating cycles. Parts were "retired for time" when 1 of 1,000 parts could potentially develop a critical fatigue crack., All 1,000 parts were retired to eliminate the possibility of catastrophic failure in flight. The tremendous cost associated with spare parts procurement under this maintenance philosophy however, generated and motivated the pursuit of the RFC development.

In 1978-79, it became clear that an impeding "Engine Spare Parts" crisis was looming over the USAF, due to the higher than expected "hot cycle usage". The engine part condemnation rate under the "Retirement For time" Engine Management Philosophy would have risen to unacceptable levels. The replacement component production would have required more Cobalt than was available in the free world. The solution to the dilemma was recognized by developing inspection technology to reliably detect small (0.005" depth) fatigue cracks in the used parts. This allowed the reuse of component that had been retired for time thus implementing the RFC maintenance philosophy.

5. EVOLUTION OF THE RFC SYSTEM: USAF HISTORICAL PERSPECTIVE

In the early 1980's, the USAF Material Lab (Man Tech Division) had just funded GE Aircraft Engines to develop an eddy current inspection system, the "ECII". Because of the competitive nature of the Aircraft engine industry, the US Air Force was unable to implement the ECII on the F100 engine. The USAF made a deliberate decision to develop a common, generic, inspection system to be used on engines without regard to specific engine manufacture. In October of 1981, the Retirement For Cause/Nondestructive Evaluation (RFC/NDE) contract was awarded to System Research Laboratories, Inc. with multiple integrated subcontractors which included Aircraft and Engine Manufacturers, NDE industry, and Research Institutes (Figure 1).

The RFC inspection system has surpassed the original intended use by becoming the USAF standard fully automated eddy current inspection station (ECIS) for the ENSIP and RFC Programs at the Oklahoma City and San Antonio Air Logistics Centers. As the generic RFC system was applied to other

engine types, increased challenges such as larger inspection envelope, new complex features, and higher throughput rates dictated evolutionary changes from Version 1, interim Version 2, and the current Version 3.

Today, the USAF has 41 ECIS: 2 Ultrasonic (UT) stations, 5 Version 1 eddy current (EC) stations, 1 Version 2 eddy current station, and 33 Version 3 eddy current stations. The Oklahoma City Air Logistics Center (OC-ALC) houses 1 Version 2 and 15 Version 3 EC stations to support the inspection required for the F101-GE-102 (B1-B aircraft), F110-GE-100 (F16 aircraft), F110-GE-129 (F16) and F118-GE-100 (B2 aircraft) engines. The San Antonio Air Logistics Center (SA-ALC) owns 2 UT stations, 5 Version 1 and 18 Version 3 EC stations to support the inspection required for the F100-PW-100/220/229 (F15 and F16 aircraft) engines.

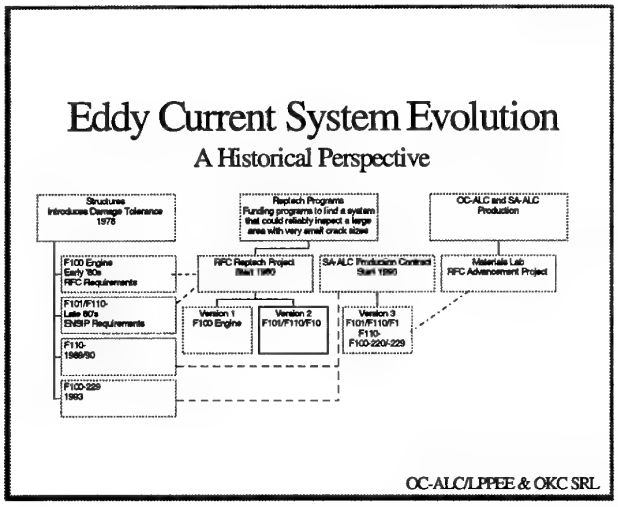


Figure 1. Eddy Current System Evolution.

6. OC-ALC RFC INSPECTION SYSTEM IMPLEMENTATION

The original inspections for these engines overhauled at OC-ALC were time consuming and very operator dependent. OC-ALC and WR-ALC/ASC funded introduction the RFC system for the inspection of F101/F110 engines. These engines automatically required advanced software and filtering to inspect the intricate component designs. The software development was further challenged by the added requirements of what is known today as "sub-geometry re-inspection." Sub-geometry re-inspection allows the inspector to inspect only the rejected holes from specified whole pattern or subset surface areas of an area pattern.

No sooner did the Version 3 station software and hardware become qualified/verified when Desert Storm began and workload surged. The station proved to be a great tool for lowering the inspection time for the engines by almost half. During the RFC-ECIS production contract delivery period, the station continued to advance with the software development of the F110-129 engine. The eddy current process is continually monitored to ensure high reliability of inspection with low maintenance cost and high throughput. As the inspection workload increases, the RFC system requires continuous evolution to respond to future needs while maintaining high throughput and low maintenance cost.

The implementation of this inspection technology at OC-ALC, has been a challenging task. Integrated teams have been formed to provide the proper infrastructure to support this effort in a quick and reliable way. The ENSIP inspection is an integral part of the entire Depot overhaul process not just a single element. The infrastructure the entire process is creates a better product.

o Abnormal reject levels act as indicator of process control problem.
o Nicks, dents, and scratches, which normally may not have been found, are being eliminated through polishing.

o Chemical and Abrasive cleaning, and Polishing techniques has been optimized for ENSIP.
o Part handling and transportation procedures were improved to minimize handling damage.
o Surface finishes have been improved on parts, that did not receive the full ENSIP inspection at manufacture, in many cases to meet the current eddy current requirements.
o Other engine component anomalies have been identified and confirmed (Figure 2).

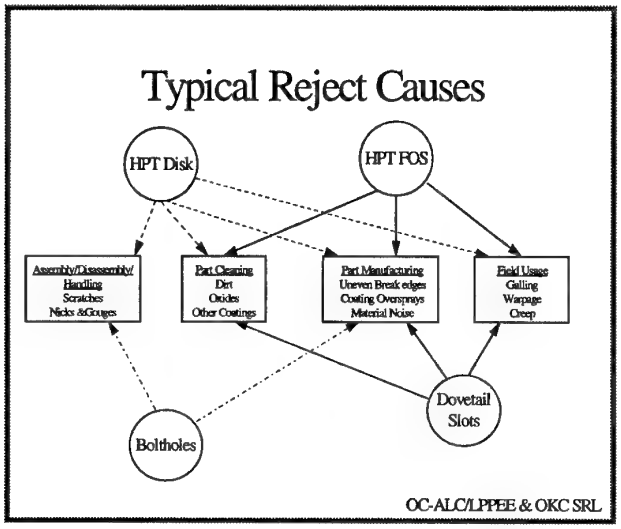


Figure 2. Typical Reject Causes.

OC-ALC will continue to pursue higher goals and expectations of the RFC system and the inspection process for the USAF and other customers.

7. RESULTS - BENEFITS

The implementation of this quantified eddy current inspection using the RFC inspection system has generated the following Benefits:

- o Increase engine availability
- o Fewer spares required
- o Decrease critical engine part failure
- o On the F100 engine:
  - -$1 billion overhaul cost savings projected
  - -6 million pounds of critical/strategic materials savings projected
  - -25:1 return on investment
- o F101 and F110 Engines have similar results as F100 engine.

8. REFERENCES

1. W.D. Rummel, G. Hardy, and T Cooper, Application of NDE Reliability to Systems., ASM Handbook Nondestructive Evaluation and Quality Control. Volume 17, May 1992 pp. 674-688.

2. "Engine Structural Integrity Program (ENSIP)", MIL-STD-1783 (USAF) 30 Nov 1984.

3. A. P. Berens, "NDE Reliability Data Analysis". ASM Handbook Nondestructive Evaluation and Quality Control. Volume 17, May 1992, pp. 689-701.

DEVELOPMENT OF RELIABLE NDI PROCEDURES FOR AIRFRAME INSPECTION

Stephen G. LaRiviere and Jeff Thompson
Boeing Commercial Airplane Group
P.O. Box 3707, Mail Stop 9U-EA
Seattle, WA 98124-2207, USA

SUMMARY

Nondestructive inspection (NDI) plays a key role in maintaining the continued airworthiness of the airplane fleet, with its ability to detect small defects with minimal disassembly. Although the responsibility for developing the inspection procedure rests with the NDI Technology engineer, collaboration with other technical communities is necessary. Structures Engineering and Customer Service representatives identify inspection requirements and provide the NDI engineers with information from fatigue tests and analysis, along with in-service issues. This collaboration has produced more than 1,000 reliable inspection procedures over the last 20 years.

LIST OF SYMBOLS

ATA	Air Transport Association
CAD	computer-aided drafting
kV	kilovoltage
mA	milliampere
NDI	nondestructive inspection
SIPD	Supplemental Inspection Planning Data
SSID	Supplemental Structural Inspection Document

INTRODUCTION

The first Boeing in-service NDI manual was developed more than 35 years ago. Today, each airplane model, from the 707 to the 777, has its own NDI manual [1]. Developing the procedures or techniques that fill these documents involves a great number of considerations. It is a task that requires

- Clear understanding of structural engineering requirements that ensure fleet safety.
- Thorough knowledge of NDI technology capabilities to ensure technical reliability.
- Complete understanding of airline customer requirements.

When expertise from these three disciplines is integrated during the NDI procedure development phase, the result is a reliable NDI system that will continue to maintain a safe fleet. Inspection economics are considered, but safety is always of paramount importance in developing reliable NDI procedures.

NDI FOR THE BOEING COMMERCIAL AIRPLANE FLEET

Beginning in the 1950s with the introduction of commercial jets, visual inspection has been the primary inspection technique. Frequent visual inspections can be rapidly and easily performed on a variety of structures. Visual inspection is particularly valuable in nondirected inspections or in those inspections in which no previous damage is suspected [2]. When fatigue tests or in-service experience indicate that a directed structural inspection is required, instrumented NDI techniques become valuable since they can detect smaller cracks and require only minimal disassembly. (See Fig. 1.)

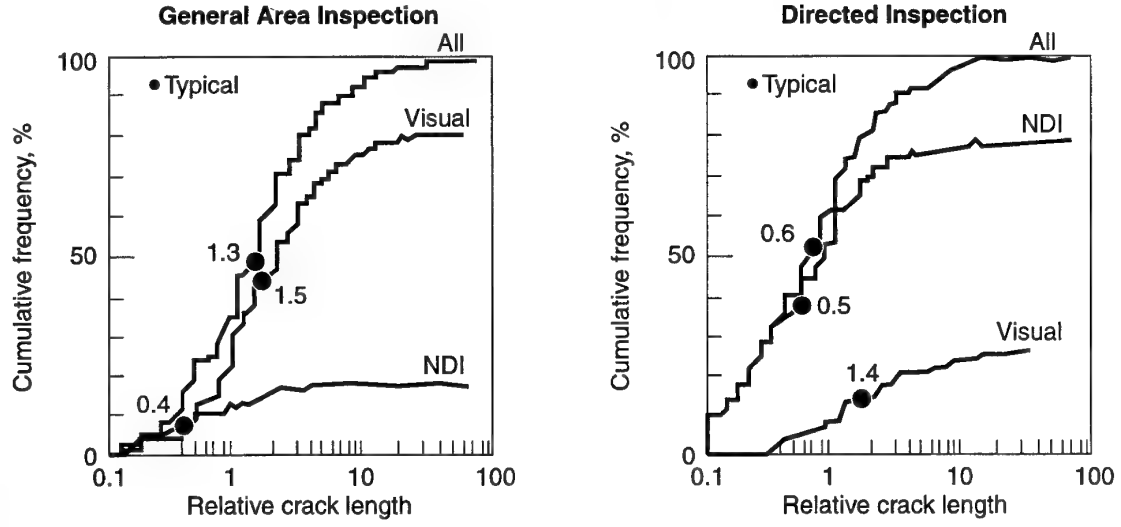


Figure 1. Distribution of Cracks Found in Service

Paper presented at the RTO AVT Workshop on "Airframe Inspection Reliability under Field/Depot Conditions", held in Brussels, Belgium, 13-14 May 1998, and published in RTO MP-10.

Today, airlines use the five major NDI techniques (magnetic particle, liquid penetrant, ultrasonics, eddy current, and radiography) with new techniques, such as thermography and shearography, becoming popular. The NDI manuals that Boeing produces for its customers are used to support

- Airworthiness directives (AD).
- Service bulletins (SB).
- Fleet monitoring programs, such as Supplemental Inspection Planning Data (SIPD).
- The Corrosion Control Program.
- Assorted service damage detection techniques, such as fire damage or composite repairs.

In the past 20 years Boeing has produced 1,149 NDI procedures, as depicted in Figure 2.

	707	727	737	747	757	767	777	Total
SB/AD	112	67	51	109	6	24	0	369
SSID/SIPD	127	83	69	141	2	0	0	422
General	41	52	60	65	60	45	35	358
Totals	280	202	180	315	68	69	35	**1,149**

Figure 2. NDI Procedure by Usage

For the 707, X-ray inspections were used extensively to ensure continued safety. The use of eddy current inspections for airplane structure was still in its infancy. Since that time, this technique has matured with the advent of shielded pencil probes and low-frequency eddy current techniques that detect small cracks. Eddy current inspections have continuously replaced X-ray inspections, as shown in Figure 3. A summary of NDI techniques with typical applications and detectable defect sizes for in-service airplane inspection is shown in Figure 4.

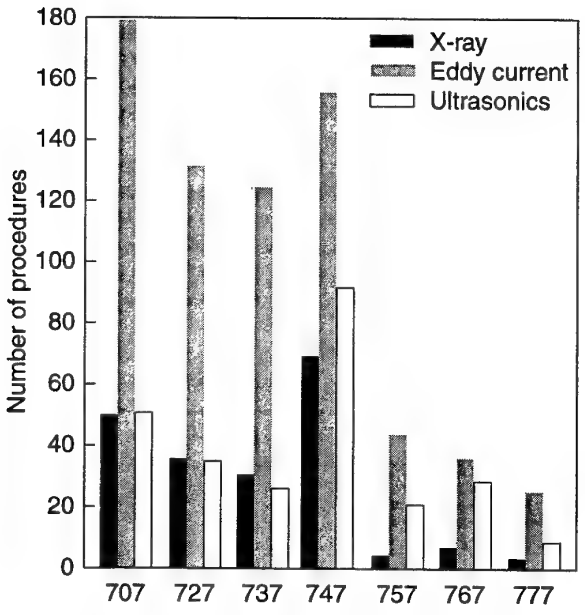


Figure 3. Inspection Method Versus Airplane Model

NDI SYSTEM RELIABILITY

A reliable NDI system combines the following elements: inspection techniques, NDI equipment, and a qualified inspector. Failure to provide proper attention to any of these elements can result in a compromised inspection system.

The first element is the NDI procedure. It provides the inspector with detailed instructions that describe how to perform an NDI inspection on a particular type of structure. At Boeing, procedures are written to comply with Air Transport Association (ATA) Specification 100. The procedures contain concise instructions that describe

- The purpose of the inspection.
- Minimum equipment requirements (including reference standards).
- Inspection parameters.
- Interpretation of results.

The steps involved in developing these procedures are discussed in the section entitled "NDI Procedure Development."

The second element of a successful NDI system is the NDI equipment. A reference standard is developed for each procedure, and it is used to define the NDI equipment requirements. Any NDI equipment that can resolve the required flaw size with the proper signal-to-noise ratio (typically 3:1) is allowed for the inspection. It is the responsibility of the operator to ensure that the equipment is operating to the manufacturer's specifications. By qualifying NDI equipment based on the reference standard, the operator is free to use NDI equipment from any manufacturer that meets the inspection sensitivity requirements specified in the NDI procedure.

The third, and potentially most important element, is the inspector. The inspector must not only understand the proper operation of equipment but must also have in-depth knowledge of the NDI technology and its limitations. With detailed knowledge of the structure, including inspection history and failure mechanisms, proper analysis of signals is ensured. Although procedures are verified prior to release, details can be overlooked. Since a knowledgeable inspector is free to identify improvements in technique to Boeing, the inspector is a valuable component of the overall NDI system reliability.

INTERNAL ROLES

Within Boeing Commercial Airplane Group in the Puget Sound area, expertise from three technical communities is combined to ensure the relevance and reliability of our NDI procedures. The three represented communities are Structures Engineering, NDI Technology, and Customer Service representatives. (See Fig. 5.)

Structures Engineering, through many hours of fatigue testing and analysis, determines fleet leading items that may require directed NDI procedures. The structures community provides an understanding of crack propagation rates, crack orientations, and failure mechanisms. All these factors are critical in developing reliable procedures. The structures community also helps establish conservative, repeat inspection intervals to ensure continued airworthiness. This is intended to allow three inspection opportunities before a crack becomes critical. These criteria may be modified for rapidly growing cracks.

Method	Material type	Defect type	Estimated minimum practicable detectable size*	Advantages	Disadvantages
X-ray **X-ray**	Metals Nonmetals	Surface, subsurface, and internal cracks (multilayered structure)	Dependent on geometry and material parameters	1. Record of test results 2. Inspects all layers of multilayered structure 3. Minimum preparation of structure in most cases* 4. Good indication of crack location and length *Airplane defueling may be required.	1. Inspection is directional for crack detection 2. Personnel evacuation from airplane during X-ray exposure 3. Defueling required for crack detection in fuel areas
Ultrasonic (pulse-echo) **UT**	Metals Some nonmetals	Surface and subsurface cracks (first layer only)	0.1 in at fastener holes or similar specifications 0.15-in general structure	1. Surface and subsurface cracks in first layer 2. Minimal airplane or part preparation 3. Access required from only one side 4. Small cracks detectable 5. Minimal inspection time	1. High operator skill 2. No record of crack indications 3. Surface contact for part being tested 4. Limited to upper member 5. Inspection is directional for crack detection
High-frequency eddy current–surface inspection **HFEC**	Metals (magnetic and nonmagnetic)	Surface cracks in Al, Ti, steel Near surface cracks (0.005 in) Al, Ti	0.030-in corner crack in holes fastener 0.1 in around fastener ends 0.2 in general surface inspection	1. Rapid inspection 2. Nondirectional 3. No paint removal, adaptable to most surface geometry	1. Careful inspection required 2. Inspection is sensitive to surface-to-probe orientation 3. Sealant removal generally required at inspection surface 4. Contact required with part surface
Low-frequency eddy current **LFEC**	Metals (nonmagnetic or low permeability)	Subsurface cracks 0.5 in below surface	Dependent on geometry and material parameters	1. Rapid inspection 2. Minimal airplane preparation 3. Second layer (within thickness penetration limit)	1. High operator skill 2. Careful inspection required 3. Significant interference from structure variables 4. Access to part surface required
Magnetic particle **MT**	Steel (magnetic) Stainless steels	Surface and near surface cracks	0.1-in-long surface crack 0.050-in corner crack with fastener or pin removed	1. High sensitivity 2. High accuracy	1. Directional 2. Visual contact with part 3. Surface finish removal desirable
Penetrant **PT**	Metals	Surface cracks	0.15-in-long surface crack 0.050-in corner crack with fastener or pin removed	1. Easy to perform 2. Minimal inspector skill required	1. Only cracks open to surface 2. Visual contact with part 3. Careful surface preparation 4. Etching required after smear metal operation
Low-frequency eddy current **LFEC**	Metals	Faying surface and second layer corrosion	10% material loss	1. Rapid inspection 2. No disassembly required	1. High operator skill 2. Careful inspection required
Ultrasonic mechanical impedance bond tester (low frequency)	Metallic and nonmetallic honeycomb structure	Skin-to-core disbond	1.0-in diameter	1. Rapid/reliable inspection 2. No couplant 3. Single-side inspection 4. Minimal airplane preparation	1. Mainly for near-side disbond only 2. Metallic maximum facesheet thickness over core: 0.10 in 3. Nonmetallic maximum facesheet thickness over core: 0.128 in 4. High operator skill
Ultrasonic resonance bond tester (high frequency)	Nonmetallic	Interply delamination	0.375-in diameter	1. Rapid/reliable inspection 2. Single-side inspection 3. Minimal airplane preparation 4. Maximum thickness: 0.438 in	1. Couplant required 2. Not conducive to large-area inspection 3. High operator skill

*Smaller defects may be detectable in specific instances

Figure 4. Inspection Methods – NDI Damage Detection

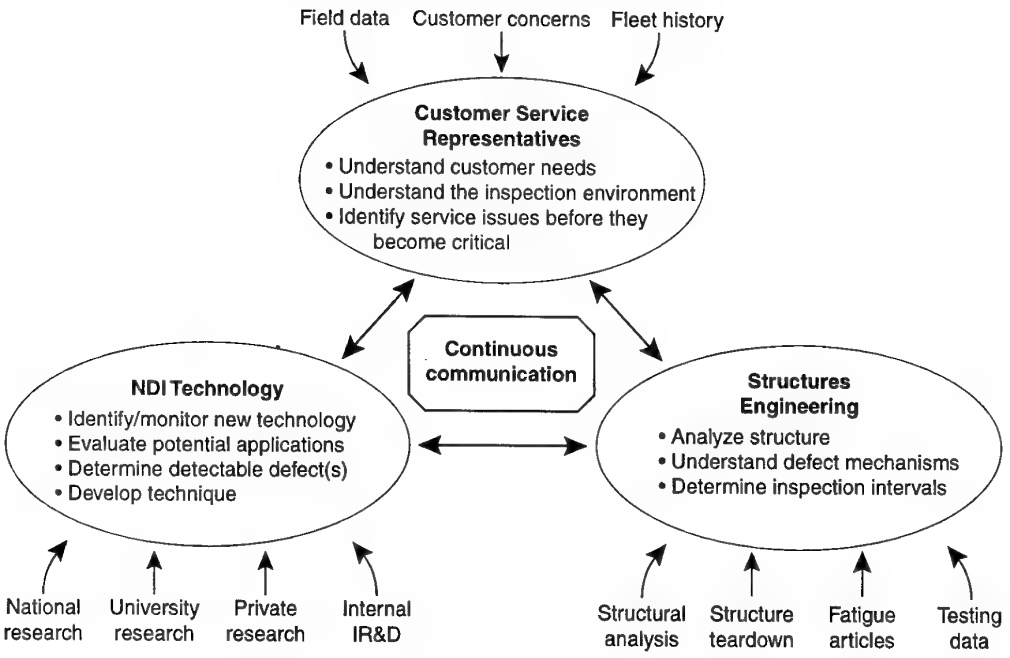


Figure 5. NDI Procedure Development Roles

The Customer Service representatives perform a number of important roles in the development of reliable NDI procedures. Through close contact with its customer airlines, Boeing is continually informed of service-related issues. By monitoring fleetwide service issues for all airplane models, Boeing Service Bulletins can be released, where appropriate, to alert all operators to potential structural inspections. The service bulletins can ultimately result in an FAA airworthiness directive. In this role, Customer Service representatives and structures engineers work together to refine their predictive models based on in-service data.

Finally, Customer Service representatives have an intimate knowledge of the customer airlines' inspection concerns and limitations. Here, safety, cost, and schedule issues are brought into the development equation.

The last organization that is involved is the NDI Technology group, which is responsible for preparing the written technique or procedure. The NDI engineers must fully understand the inspection goals for each structural component. By understanding how defects propagate in a structure, they can assess a variety of NDI technologies available in their "NDI toolbox" to develop a reliable, cost-effective inspection procedure. It is imperative that the NDI engineer

- Communicate with Structures Engineering and Customer Service representatives.
- Understand all available NDI technologies in detail (including visual inspection and its limitations).
- Possess the ability to develop a written procedure that clearly describes the steps needed to perform the inspections.

The NDI engineer has another very important role: to research and implement new and improving technologies and to develop new inspection techniques that can be made available in the NDI toolbox for future use.

In many cases, there is an immediate need for an inspection procedure when an inspection requirement is identified. Prior research ensures that procedure development time is minimized by selecting the proper tool from the wide variety of reliable, proven tools in the NDI toolbox. Rapid NDI procedure deployment is further enhanced with a continuous dialogue among the three technical communities. Teamwork brings the best results.

The following is a hypothetical situation that illustrates the importance of close communication. Due to a service problem, the customer representative requires an analysis of a particular component. After review and analysis, the structures engineer determines that a repeat inspection interval of 3,000 cycles is required and an NDI procedure must be developed to find a 0.1-in crack in buried structure.

The task is then given to the NDI engineer. After performing laboratory experiments and visiting the actual structure, the engineer concludes that the NDI technology can detect only a 0.2-in crack to give a repeatable 3:1 signal-to-noise ratio. The structures engineer recalculates and finds a 1,500-cycle inspection interval using a 0.2-in crack. If a crack exists, there will be three opportunities to inspect and identify it before the crack becomes critical. The Customer Service representative feels that this inspection cycle fits typical airline maintenance intervals, and the NDI engineer is able to complete development of the inspection procedure by including all the required data. The next section will address the procedure development details.

NDI PROCEDURE DEVELOPMENT

For the sake of discussion, an area of an airplane has been identified by the Customer Service representative or Structures Engineering, and a nondestructive inspection procedure has been requested. The following steps are taken by the NDI engineer to develop a reliable NDI procedure that will find the required defects while minimizing false calls.

Research Inspection Parameters
The first step is to understand the inspection parameters:

- Material (alloy, conductivity, permeability).
- Structural geometry (thickness, edge boundaries, stack-up, fastener spacing).
- Accessibility.

These parameters, the defect type (stress corrosion cracking, fatigue cracking, or disbonding), and the desired detectable defect sizes will drive the selection of the optimum inspection methodology. Alternative inspection methodology may be required when certain inspection parameters are present that are known to cause inspection difficulties. For instance, a steel structure may not lend itself well to an eddy current inspection because of the variations in permeability in the steel. This may increase the false-call rate. Yet, if an ultrasonic inspection is used, special attention to grain orientation may be required. Appropriate selection of the best NDI method is made by experienced NDI engineers using lessons learned in the development of procedures for similar structures.

Assess Inspection Options
The next step is to review the previously mentioned data, obtain available fatigue-test or in-service structure with "real defects," and physically go to an airplane to determine access. It is now possible to review candidate inspection options. A knowledgeable NDI engineer is invaluable during this step to quickly determine the most appropriate NDI method.

In the process of assessing options, real structure may not be available for testing. This necessitates simulating the structure using a mockup to aid in laboratory development. Instances where this is valuable are illustrated in the following examples. First, the NDI engineer may need to mock up lower edge margins on a subsurface eddy current inspection. As the frequency is reduced and gain is increased, the sensitivity to lower edge margins may increase false calls. The influence of the edge margin extremes must be understood when the procedure is developed.

Another example is that of a lug inspection. Mockups of lugs with cracks at various angles may be needed to assess the limitations of an ultrasonic inspection for a given geometry. The mockup structure typically lays the foundation for the reference standard that ultimately will appear in the inspection procedure.

The power of computer modeling helps accelerate this part of the process. Computer-aided drafting (CAD) systems are routinely used to design reference standards. By reducing development time and maximizing the ultrasonic signal, computer modeling is also quickly becoming an important tool for designing ultrasonic positioners. As more NDI models are developed and interfaced with CAD files, additional time savings will be achieved.

Develop Inspection Parameters
At this point, the inspection challenge is understood and various mockups are designed. Next, inspection parameters need to be defined. In the case of an X-ray inspection, these may include kilovoltage (kV), milliamperes (mA), film type, shielding, and penetrameters. Eddy current inspection will require frequency, lift-off, probe type, and equipment considerations. And finally, in the case of ultrasonic inspections, parameters such as frequency, sensor diameter, positioning fixture design, and filters are determined.

At this time the NDI engineer communicates the actual detectable defect size to the structural engineer, and the reference standard is finalized with the help of the mockup. If the detectable defect size is larger than the required size, the inspection interval may be adjusted. It should be noted that many times smaller defects can be detected, yet the NDI engineer develops the procedure with the larger size to improve reliability. In the process of determining the system performance, a 3:1 signal-to-noise ratio is used. This ensures that a very distinguishable defect will be clearly identifiable above the background noise.

Write Procedure
The least exciting segment of the process for the NDI engineer is the task of writing the procedure. Clarity of the procedure is very important. Therefore, Boeing NDI engineers write procedures using "Simplified English" and follow the guidelines of ATA Specification 100, which requires these sections:

- Purpose of inspection.
- Equipment required.
- Preparation and cleaning.
- Equipment calibration.
- Inspection procedure.
- Inspection results.
- Acceptance/rejection criteria.

Verify Procedure
To ensure that the airline inspectors will be able to implement the procedures, all procedures are verified on actual airplanes. Although care is taken throughout the process to eliminate the unknowns, this final step better ensures proper performance of the procedure. Although procedure verification is sometimes difficult or costly, it remains a very important step and is not overlooked.

CONCLUSIONS

To develop reliable inspection procedures, cooperation and continuous communication is required between Structures Engineering, Customer Services representatives, and NDI Technology engineers. This communication triangle allows inspection strategies to be quickly modified to fit specific structural inspection requirements based on minimum defect size, appropriate inspection interval, and best available NDI equipment. The resulting NDI procedure, in the hands of a skilled inspector, will ensure the continued safety of the airplane fleet. As new equipment is designed and new inspection methodologies are developed and tested, they are added to the NDI toolbox, with an ever-present goal of reducing procedure development time, increasing inspection capabilities, and improving overall NDI system reliability.

REFERENCES

1. "Boeing Nondestructive Test Manuals," Boeing Commercial Airplane Group, Seattle, Washington.
2. Goranson, U. G., "Damage Tolerance Facts and Fiction," in "14th Plantema Memorial Lecture," August, 1993, p. 13.

PROBABILITY OF DETECTION OF CORROSION IN AIRCRAFT STRUCTURES

J. P. Komorowski
D. S. Forsyth
D. L. Simpson
R. W. Gould
Institute for Aerospace Research
National Research Council
Building M14, Montreal Road, Ottawa ON Canada K1A 0R6
Email: jerzy.komorowski@nrc.ca

SUMMARY

High cost and safety concerns related to aircraft corrosion indicate the need for changes to the current "find-it-fix-it" philosophy for corrosion management. Developments in non-destructive inspection techniques will lead to multidimensional corrosion metrics to support the corrosion damage assessment of structures. Data fusion techniques are proposed to aid in the interpretation of the multiple non-destructive inspections typically required for corrosion damage quantification. Corrosion reliability in terms of probability of detection (POD) is proposed as a requirement for safety related corrosion detection. Quantification of the POD for field corrosion inspections is limited by the subjective manner in which detected corrosion is characterised. Corrosion metrics need to be identified to provide consistency to characterisation of detected corrosion, to provide input to corrosion analytical assessments and to provide the basis for POD evaluations.

LIST OF SYMBOLS

DT - damage tolerant
FAA - Federal Aviation Administration
IAR - Institute for Aerospace Research
POD - probability of detection
PoFA - probability of false alarms
PSE - principal structural element
USAF - United States Air Force
SRM – Structural Repair Manual
SB – Safety Bulletin

1. INTRODUCTION

Annually, the corrosion of metals costs the United States economy nearly $300 billion or 4% of the GNP (Ref. 1). More specifically, the annual direct cost of metallic corrosion on aircraft in the US is approximately $13 billion (Ref. 2). In the United Kingdom (UK), the estimate is that the fight against corrosion costs around 4% of the annual UK gross national product (Ref. 3). A similar situation exists in other NATO member nations in both the military and civil sector.

Corrosion costs the US Air Force more than $1 billion annually (Ref. 4). A good example is the increase in cost of maintenance of the KC-135 fleet of tankers. The US Air Force cited the age of the aircraft and a lack of replacement parts as primary reasons for the increased maintenance time. However, other contributing factors included the lack of information about the condition of aircraft coming into the depot, and additional work required to detect, repair, and prevent corrosion.

Improvements in corrosion detection are expected to have significant impact in lowering the high costs of corrosion management. Over the last 10 years, the efforts to develop methods of corrosion detection in aircraft structures have produced a number of new or improved NDI techniques. D Sight – an enhanced visual method, thermography, and pulsed eddy current are examples of these new techniques. In the United States the Federal Aviation Administration (FAA) has sponsored several of these developments following the well-publicised Aloha Airlines accident (Ref. 5). In spite of these extensive efforts, the FAA program has not led to any of these new corrosion detection techniques being adopted by the airlines. The reasons for this include: cost-benefit of these techniques not fully accepted by the operators or OEM's; corrosion is an accepted economic problem, not a safety problem therefore regulatory requirements are not precise; and, there is no real quantification of improved inspection performance. Corrosion detection techniques have not been rigorously evaluated using a statistically valid probability of detection (POD) approach. POD numbers do not exist for the current most common approach – visual inspection followed by manual single-frequency eddy current inspection.

The only study to date, which attempted to compare NDI methods for corrosion detection, was sponsored by the United States Air Force in support of its KC-135 fleet (Ref. 6). In this study, the POD approach developed for surface crack detection was used to compare three eddy current procedures. However, to reduce the cost of the study, the test samples used were lap joints with the thickness of the first layer at the faying surface reduced using EDM. This was a significant drawback since these flaws had sharply defined edges and surface roughness different from corroded sheets in which pitting, exfoliation and intergranular corrosion are often present simultaneously. These samples also did not exhibit pillowing, a plastic deformation of the joint between rivets caused by the corrosion product.

To evaluate various NDI techniques for corrosion detection, the FAA has tasked the Aging Aircraft NDI Validation Center (AANC) to develop a corrosion detection experiment (Ref. 7). The main focus of this experiment is the detection of hidden corrosion in lap joints with emphasis on 5 to 10% thinning. Other issues such as corrosion type, pitting, stress redistribution and surface morphology are not addressed.

A great deal of effort in the AANC and USAF studies has been expended on designing the experimental procedures in an

Paper presented at the RTO AVT Workshop on "Airframe Inspection Reliability under Field/Depot Conditions", held in Brussels, Belgium, 13-14 May 1998, and published in RTO MP-10.

attempt to provide objective, quantitative, and systematic evaluation of the reliability of an NDI process. In the AANC task, various sources of test panels were being considered to achieve realistic conditions and statistically valid results. The AANC will attempt to generate POD and probability of false alarms (PoFA) for the corrosion techniques evaluated. However, the referenced paper does not describe how these figures will be calculated.

Both the USAF and AANC POD studies quantify corrosion in terms of area and percent thinning. The percent thinning approach ignores at least two important factors: first - corroded surface morphology depends on absolute thickness loss and cladding thickness; and second – sheet thickness tolerance may affect corrosion loss estimates expressed in terms of percent thinning. In this paper the various forms of corrosion and their impact on structural integrity and possible maintenance options will be considered in formulating the requirements for developing a POD approach to corrosion NDI.

2. THE NEED FOR RELIABILITY CHARACTERIZATION OF CORROSION NDI

In parallel with the efforts to develop better corrosion detection techniques, several programs currently in progress in Canada at the NRCC and in the US under USAF sponsorship are attempting to develop analytical capabilities for corrosion damage assessment in aircraft structures. These analytical models will require, as input, quantified corrosion damage data from NDI. The outputs from the corrosion damage assessments will allow the planning of maintenance actions depending on the current state of corrosion, its influence on structural integrity and its projected growth.

Boeing has already, in a limited way, implemented this approach. Some service bulletins allow operators to continue operating the aircraft with corrosion in lap joints provided that it is less than 10% of the original sheet thickness. Frequent re-inspections are mandated to ensure that the 10% limit is not exceeded (Ref. 8). The intent of such programs is to move away from the current "find it-fix it" philosophy of dealing with corrosion to managed proactive maintenance that considers both safety requirements and economic issues. One suggested approach for the future is "find it sooner-evaluate-plan-fix". Advances in NDI allow earlier identification of corrosion. The impact of corrosion damage on structural integrity is being studied by many programs with the intent of defining a corrosion damage framework for determining residual life and residual strength. This framework can be used to plan maintenance of the corrosion such that down time and repair costs are minimised and the risk of incurring additional damage is reduced. There is substantial evidence that suggests that fixing corrosion often results in extensive damage (Ref. 9).

Recent studies in transport aircraft fuselage lap joints indicate that corrosion may not only be an economic issue but a safety issue as well (Ref. 10). Corrosion at faying surfaces of riveted sheets produces the well-known pillowing effect. It has been shown that pillowing can result in stress levels exceeding yield at corrosion sheet thinning as low as 5% and to the shifting of the stress critical location established for corrosion-free joints. These high pillowing stresses in the presence of a corrodent also lead to high-aspect ratio cracks. Pillowing cracks are rarely detected because current inspections have been set up to detect surface-breaking cracks in fatigue critical rivet rows. Limited fractographic studies done at the NRCC have shown that some of these cracks grow under fatigue loading (see Figure 1). While more research is needed to quantify the effects of these cracks on residual strength and residual life of fuselage joints, the interaction between corrosion and fatigue raises concerns for the safety of continued operation of corroded fuselages.

A recent sample of corrosion problems in three aircraft fleets in the USAF has uncovered corrosion damage which raised safety concerns (Ref. 11). The major conclusions of this study were:

- Corrosion damage in critical structure has resulted in an initial flaw size that dramatically reduced the predicted life of a critical component.
- Corrosion damage in structural elements has changed a component's status from non-critical to critical.
- Durability Assessment and Damage Tolerance Analysis (DADTA) studies do not provide inspection locations and intervals for the observed corrosion problems.
- Corrosion damage has led to flight safety concerns for a number of locations on the C-5 and prompted changes to inspection procedures.
- Corrosion has led to premature cracking of major structural elements, significantly earlier than DADTA predictions.

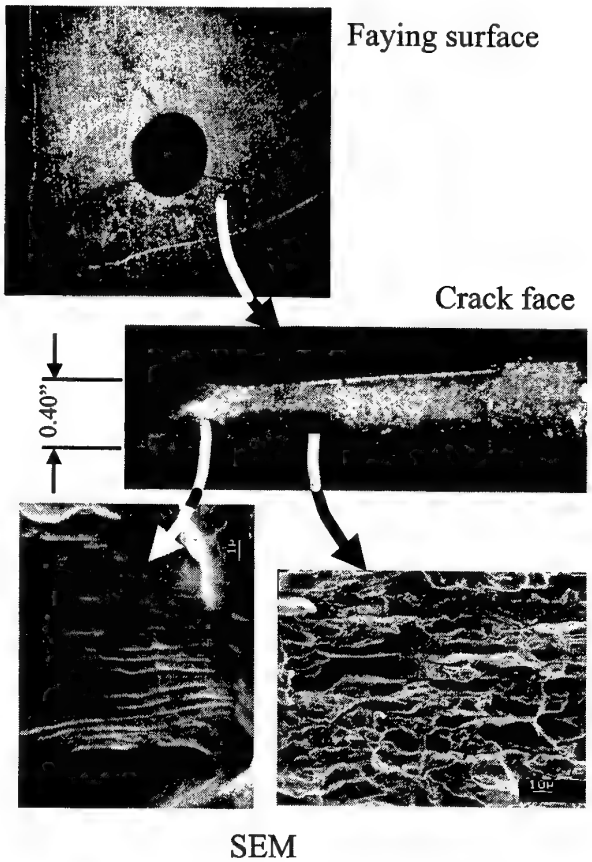


Figure 1. Section of a skin from a Boeing B727-90C manufactured in 1966. The skin was removed at approximately 57,000 cycles and 72,000 hours. The skin ran between BS 360-440 and STR 19-26L. The cracks occurred at BS 440, two rivet rows below STR 24L in the first layer. Cracks are visible on the faying surface (right). In the middle is the crack face after the crack was opened. On the left are two SEM images, the upper from the crack front shows fatigue striations, the lower shows corroded crack face (Ref. 10).

From the above studies it seems that the main reason why corrosion may become a safety issue is the fact that it may lead to cracking in the areas which are not subject to crack inspections. Very low levels of corrosion may shift the critical cracking to another area. Consistent with the DT approach, this requires that methods of corrosion detection with known reliability (POD) be applied to maintain safety.

3. TYPES OF CORROSION

Corrosion in aircraft may appear in various forms depending on the alloy, product form, corrodent, general conditions and residual stress (ref. 12). This complicates the metrics of corrosion and therefore also complicates the quantification of detection reliability. The following list was adopted from (ref. 2) which lists six types of corrosion.

3.1 Galvanic Corrosion

Galvanic corrosion is a very common form of corrosion that results from contact between dissimilar metals. A difference in the electrode potential of the two metals and the difference in the surface area of the dissimilar metals drive the process. Galvanic corrosion is responsible for much of the corrosion in aircraft.

3.2 Pitting

Pitting is another form of corrosion that results when the anodic site in the electrochemical reaction corresponds to a local microstructural discontinuity, such as an inclusion, grain boundary, or even a scratch, on an otherwise large cathodic surface area.

3.3 Crevice Corrosion

Crevice corrosion is a form of localised corrosion that occurs near an area of a metal surface adjacent to another metal that is sheltered from full exposure to the environment. The reaction between the oxygen in the crevice and the rest of the metal causes a gradient in the oxygen concentration, and thus a difference in electrode potentials and a flow of current.

3.4 Intergranular Corrosion

Intergranular corrosion occurs at or adjacent to the grain boundaries of a metal or alloy. The actual mechanism of the corrosion varies with metal system. This attack at the grain boundaries can cause entire metal grains to become dislodged. Leakage of corrosive fluids, loss of effective cross sectional area, and mechanical failure can result.

3.5 Erosion-Corrosion

Erosion-corrosion, as its name suggests, results from the actions of corrosion and erosion in the presence of a moving corrosive fluid, causing accelerated loss of the metal.

3.6 Hydrogen Induced Environmental Cracking

Hydrogen induced environmental cracking results from the combined action of a tensile stress and a corrosion reaction that leads to the production of nascent hydrogen at the cathode. This form of corrosion often causes an otherwise ductile metal to fail. Failures resulting from environmental cracking are often disastrous because they occur in metals that usually have good corrosion resistance.

3.7 Stress Corrosion Cracking

Stress Corrosion Cracking, similar to hydrogen induced environmental cracking, results from the combined effects of a tensile stress and a specific environment. It may lead to anodic dissolution of grain boundaries or specific crystallographic planes.

4. CLASSIFICATION USED TO EVALUATE CORROSION ON AIRCRAFT

Once corrosion of any type is detected, an effort must be made to quantify the damage for both reporting and repair purposes. In the absence of generally accepted quantitative tools for evaluating the structural integrity implication of corrosion damage, a subjective assessment of detected corrosion levels has evolved as the basis for airframe maintenance.

There are currently at least two common classification systems. One is most often associated with military operators (Ref. 13), however, the same system is mentioned by Boeing in a Structural Repair Manual (Ref. 14). This system classifies corrosion as light, moderate, or severe based mostly on visual appearance and depth of attack or material loss:

LIGHT: Characterised by discoloration or pitting to a depth of approximately 0.001 inch (0.025 mm) maximum - this type of damage will normally be removed by light hand sanding and a minimum of chemical treatment.

MODERATE: Appears similar to light corrosion except there may be some blisters or evidence of scaling or flaking. Pitting depth may be as much as 0.01 inch (0.25 mm), in which case the damage should be removed by extensive hand sanding or mechanical sanding.

SEVERE: General appearance may be similar to moderate corrosion with severe blistering exfoliation and scaling and flaking. Pitting depths will be deeper than 0.01 inch (0.25 mm) - removal of this type of damage normally requires extensive mechanical sanding or grinding.

This system is used for corrosion evaluation after the initial inspection and cleaning.

The commercial airline industry has attempted to quantify corrosion as a means of determining effectiveness of corrosion prevention and control programs. The FAA (Ref. 15) and Boeing (Ref. 16) divide corrosion into three levels:

Level 1 Corrosion

- Corrosion damage occurring between successive inspections that is local and can be reworked / blended-out within allowable limits as defined by the manufacturer (e.g., SRM, SB, etc.)

or

- Corrosion damage that is local and exceeds allowable limits but can be attributed to an event not typical of the operator's usage of other aeroplanes in the same fleet (e.g., Mercury spill)

or

- Operator experience over several years has demonstrated only light corrosion between successive inspections but the (results of the) latest inspection and cumulative blend-outs now exceed allowable limit

Level 2 Corrosion

- Corrosion occurring between successive inspections that requires rework / blend-out that exceeds allowable limits, requiring a repair or complete or partial replacement of a

principal structural element (PSE) as defined by the original equipment manufacturer's structural repair manual

or

- Corrosion occurring between successive inspections that is widespread and requires blend-out approaching the allowable rework limits

Level 3 Corrosion

- Corrosion found during the first or subsequent inspections that is determined (normally by the operator) to be a potentially urgent airworthiness concern requiring expeditious action.

An effective program is one that controls corrosion of all primary structure to level 1 or better.

The FAA ageing aeroplane corrosion programs include a mandatory system for reporting Levels 2 and 3 corrosion findings to the manufacturers. This system is based on allowable rework/blend-out limits based on suggested thickness loss, area affected, whether or not it is "wide-spread" corrosion, and the type of structure affected (PSE or other). Expected rates of corrosion based on the operator's experience are also a factor in how the corrosion is rated.

Fundamentally, however, both the above systems are very general, subjective and open to individual interpretation. They do not provide the quantitative corrosion characterisation data necessary for use in a damage tolerance assessment or for POD studies.

5. DAMAGE TOLERANT DESIGN CONSIDERATIONS FOR CORROSION DAMAGE CLASSIFICATION

The damage tolerant (DT) design philosophy is founded on the quantification of crack growth and reliable inspection procedures. Structure designers assume that cracks of certain minimum size already exist at time of entry to service. These crack sizes are based on equivalent initial flaw studies from as-manufactured parts and manufacturer's quality control processes. Inspection onset and inspection intervals are set based on estimated crack growth rates and inspection resolution/ reliability. However, corrosion can accelerate the formation and growth of cracks, significantly reducing residual strength and life of the structure, and hence rendering the estimated inspection interval non-conservative. Corrosion can also affect the definition of critical locations and the type or shape of crack that can be detected. The high aspect ratio cracks in lap splices (Figure 1) is a prime example of this latter issue.

Currently, material thickness loss allowables have been introduced to protect the structure from overload failures. However, in some structures, deformation caused by expanding corrosion product has been shown to produce far greater effect on stress than thinning as shown in Figure 2 (Ref. 17).

Hoeppner et al. (Ref. 13 and 18) pointed out that fatigue critical damage, such as corrosion pits, is not properly accounted for in the current corrosion classification system. He noted that a potential weakness in this type of system could be, for example, that a 0.01 inch deep corrosion pit may represent a severe risk of fatigue cracking in one material, but not in another, yet the material thickness loss allowable for both materials may not be violated. Also, there could be multiple pits and embrittlement resulting from the pitting. Perez (Ref. 19), Doerfler (Ref. 20) and others have used an equivalent initial flaw size (EIF) approach of accounting for pitting for purposes of structural analysis. Bush et al. (Ref. 21) have shown that the EIF approach may also be used to assess the effect of intergranular corrosion. Brooks (Ref. 2 – December 1997 TIM meeting) proposed an approach which would account for pitting and corrosion surface morphology through modification of stress intensity factors.

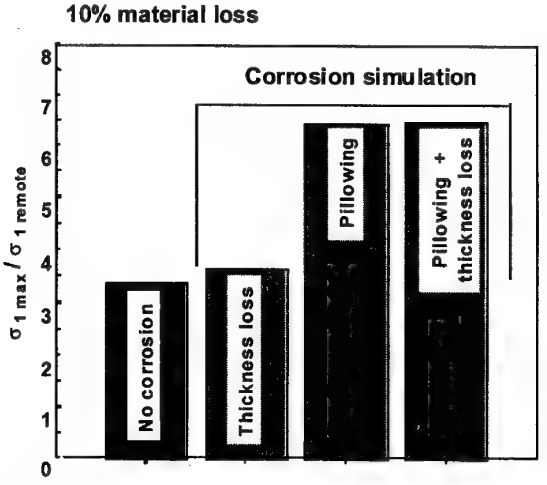


Figure 2. Effect of corrosion pillowing in a lap joint on maximum stress as compared to thinning due to corrosion (Ref. 17).

If DT analysis is to account for both cracking and corrosion then structural analysis models are required that account not only for stress modification due to thinning, but also for corrosion-related effects such as pillowing, pitting and surface morphology. These enhanced models will have to be applied along with fracture mechanics models. A fundamental requirement will be an ability to quantify real corrosion in terms that can be used as input to these models. POD data from NDI methods is required to characterise safety risks. In-service corrosion rates can also have an effect and must be considered in establishing inspection intervals.

Corrosion damage is complex and its effect ranges from changes in basic material properties to changes in the applied stress on the structural detail. It is best characterised using several NDI techniques in parallel or in sequence, where each technique is capable of quantifying one or two of the corrosion damage dimensions. NDI techniques are also complex and data fusion is one means of simplifying multimode NDI interpretation difficulties. It is postulated that a corrosion damage POD of a data fusion NDI system should be better than a single mode NDI. The proposed process of incorporating data fusion into DT assessment is shown in Figure 3.

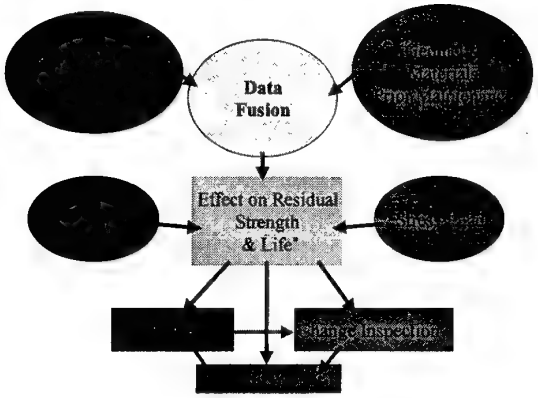


Figure 3. Assessment of structural integrity with multimode NDI and data fusion

6. NDI FOR CORROSION

A great variety of NDI techniques have been applied for detection of hidden corrosion in airframe structures. Millions of dollars have been invested in the research and development of NDI for this problem, however, both military and commercial maintenance operators still predominantly rely on visual inspections followed by manual single frequency eddy current techniques where the visual inspection isolates a potential problem. The following section will discuss some of the more accepted techniques and describe their advantages and disadvantages in terms of corrosion detection reliability. These issues must be addressed in quantifying the corrosion POD for these techniques.

6.1 Visual Inspection

Visual inspections are the original NDI technique. Commercial operators rely on visual inspections for a number of problems including hidden corrosion. In this case visual inspection depends on the effect of pillowing, and it is generally accepted that visual inspections will not find corrosion below 10% thickness loss. It should be noted that there is little experimental evidence supporting this figure. Visual inspections suffer from a number of problems, including low sensitivity, poor reliability, and poor record keeping.

6.2 Ultrasonic Techniques

Ultrasonic techniques are widely used and well developed. The application of ultrasonics to the study of hidden corrosion has been limited, due to some of the physical characteristics of the corrosion problem. The two techniques usually applied to these problems are pulse-echo and guided wave ultrasonics.

Pulse-echo ultrasonics is conceptually simple and is used in many applications. It can make highly accurate thickness measurements. Its ability to measure below the first layer in multilayer structures which may have sealant, adhesive bonding, or corrosion product between layers; is severely limited. Simple thickness measurements are also not necessarily indicative of corrosion loss: sheet tolerances in thin sheets are of the order of 5 to 8% (Table 1). This technique is also quite slow, requiring raster scanning of the entire area to be inspected. Problems of coupling the ultrasonic energy to the specimen can be overcome by using laser excitation or special probes for air-coupled ultrasonics, or squirter or dripless bubbler systems. The robotic systems developed for scanning large aircraft are very expensive, and have not been adopted by commercial operators.

Table 1. ANSI 2024 - T3 alloy sheet tolerances.

Nominal Sheet Thickness (in.)	Allowable Tolerance (in.)	% of nominal thickness
0.040	±0.0030	7.5
0.050	±0.0030	6
0.063	±0.0035	5.5
0.072	±0.0035	4.9
0.081	±0.0040	4.9

The application of guided wave ultrasonics to the problems of hidden corrosion is relatively new. They have been shown to be extremely sensitive to the sealant/bond integrity in joints, and may be able to provide estimates of thickness in the first layer. In a corroded multilayer structure, little energy will be transmitted to the deeper layers and recovered at the surface. Thus it is unlikely that much information about deeper layers can be derived from this technique. Benefits of the guided wave technique are in speed and ease of application.

6.3 Eddy Current Techniques

Eddy current techniques are another common set of NDI techniques. Techniques used for hidden corrosion include single frequency, multiple frequency, and pulsed eddy current. Any of these techniques can be performed manually or robotically. Automated interpretation of these techniques is being developed.

Single frequency eddy current data are suitable for measuring the thickness of the first layer of a multi-layer metallic structure. Between layers, sealants, adhesives, corrosion product, or air gaps may exist in any combination. These factors prevent the single frequency eddy current technique from determining with any accuracy the thicknesses of deeper layers. This technique is slow, as it requires raster scanning of the entire area under inspection. Scanning can be automated, but this is expensive. The greatest benefit of this technique is in its familiarity and ease of use. The reliability of this method for the detection of material loss is not well demonstrated. A USAF-sponsored study (Ref. 6) carried out on simulated aircraft lap joint coupons found that this technique could not detect 10% material loss at a probability of detection of 90%.

Multiple frequency eddy current techniques have been developed in response to the limitations of traditional eddy current techniques. These techniques use familiar instruments and probes, but employ two or more frequencies to excite the probe. Frequencies are chosen to maximize sensitivity to first and second layer thicknesses, and minimize lift-off and other extraneous factors. These techniques provide more information than single frequency eddy current, but are affected by the same factors. While more sensitive than single frequency techniques, the previously mentioned USAF study (Ref. 6) found that of two dual frequency and one multiple frequency eddy current technique tested, none could detect 10% material loss at a probability of detection of 90% in simulated aircraft lap joint coupons. Thus quantification of corrosion by these methods is unlikely.

A recent development in eddy current NDI is the use of pulsed excitation instead of continuous wave excitation. This should provide more information than any other eddy current technique, and experimental results on simulated corrosion show promise for good sensitivity and the ability to quantify material loss. However, the authors are not aware of any

published data available on the application of pulsed eddy current to specimens which have corrosion from service. This technique is still not well understood, although the instrumentation is similar to traditional eddy current, interpretation of results is quite different. It is possible to use automated scanning, but this technique is slow due to the necessity to do raster scanning and maintain contact between probe and specimen. At this time, pulsed eddy current techniques show great promise for corrosion identification and quantification, but demonstration on actual corroded aircraft joints has not been performed.

6.4 Enhanced Visual Techniques

Recently, developments have been made in enhanced visual NDI techniques, specifically D Sight and Edge of Light (EOL). Both of these techniques rely on measurements of corrosion pillowing which can be used to infer total material loss. Both are faster than raster scanning techniques, and at this stage D Sight is faster than EOL. Both suffer from the fact that they are new techniques, and there is little training infrastructure or experience. D Sight may be more difficult to interpret because of variations in sensitivity across the D Sight inspection image. EOL may be able to detect cracks which extend beyond fastener heads during the same inspection used for corrosion detection. Currently, research is being performed in the development of automated interpretation of both D Sight and EOL inspections of lap splice joints for corrosion (Ref. 26, 27). In the previously mentioned USAF study (Ref. 6), D Sight was more sensitive than the eddy current techniques, but had high false call rates, due to a lack of operator training and experience with this new technique.

6.5 Other NDI Techniques

Many other NDI techniques have been used for corrosion detection in airframe structures, and it is outside the scope of this paper to describe them all. Neutron radiography or back-scattered x-ray techniques may be more sensitive than any other available technique, but they are not economically viable for depot use. They are important for verification of specimens without disassembly which is needed in technique development.

Thermography has been applied to this problem, but suffers from a number of drawbacks. Interpretation is difficult, and sensitivity to multilayer corrosion in specimens from retired aircraft has not been demonstrated. This technique could be very fast.

The magnetic-optical imager (MOI) is another technique which may be practical for the inspection of airframe structures for corrosion and cracking. It has received limited acceptance for crack detection, but its corrosion thinning detection ability is below that observed for single frequency eddy current.

7. DEVELOPMENT OF NEW CORROSION METRICS

The comparison of NDI techniques for a particular application is done through controlled evaluation of inspection data which results in probability of detection (POD) and probability of false call (PoFC) information. The relationship between the POD of a flaw and some characteristic measure of a flaw is plotted. Whatever measure is used should be significant for structural integrity: for cracks, crack length is generally used. For the detection of hidden corrosion in airframe structures, the optimum measure of corrosion is less obvious because of the various types of corrosion and types of materials.

The most relevant POD determination is done using field data since it is a measure of the "real" performance of the technique. For corrosion detection, the very subjective characterisation schemes negate this approach at this time, and therefore reliance must be placed on more controlled evaluations and round-robins. The fact that there is no optimum way of characterising corrosion is a major deterrent to using field data, even for the most common visual and eddy-current techniques. The cost and complexity of generating controlled POD studies for corrosion will necessarily lead to a severe prioritisation of what techniques, materials and structural details would most benefit from POD characterisation based on safety issues.

A composite metric for corrosion may be needed to account for the many possible structurally significant effects, including averaged material loss, pitting depth and distribution, pillowing stresses, exfoliation, stress corrosion cracking, and others. Inputs required by structural analysis models must also be considered in determining an appropriate corrosion metric. It is likely that different metrics may be used for different structures and materials.

Another important issue in determining what measures to use for corrosion NDI are repair procedures. If repairs can only be carried out on a finite size, such as the joint length between two stringers or build stations (approximately 500 mm), the NDI technique need only measure corrosion damage with enough precision to make a decision whether to repair. This observation has significant impact on the method of false call rating. If an NDI inspection identified that a 300 mm length of joint is corroded and requires repair, the minimum length (500 mm) will be opened. A post teardown inspection that indicated only a 200 mm length was corroded and required repair may lead to the conclusion that the technique produced a 30% false call rate. It is postulated that the correctly identified need to repair and open the 500 mm section is more significant than the 100 mm difference between the actual and NDI assessments.

Secondary effects may also be important. For example, the phenomenon of corrosion pillowing may also need to be considered as input into a corrosion metric for some structures. It has been shown (Ref. 17) that pillowing due to minor thickness losses can cause very high stresses. These stresses can affect residual strength and crack growth, for example causing cracks to initiate and to propagate farther without breaking the top surface of a lap joint. Thus the pillowing can change the detectability of cracking.

Although there are no accepted models for the combined influence of corrosion and fatigue effects on structural integrity, there is a growing awareness that corrosion cannot be fully characterised through a thickness loss measurement alone. As previously mentioned, thickness loss is also confounded by the normal manufactured sheet tolerances, which may easily be 4% or greater. This has important implications for NDI.

Currently used eddy current or ultrasonic techniques measure thickness loss alone. Eddy currents are a diffuse process, and spatial resolution may not be fine enough to discern pitting distributions. Ultrasonics may be sufficient for this purpose, but sensitivity is severely limited beyond the first layer of multilayer structures. Neither technique measures pillowing, which may also be structurally significant. Enhanced visual methods such as D Sight or EOL can measure pillowing, which is correlated with total material loss. They cannot distinguish the extent of corrosion on individual layers nor can they identify corrosion pitting in the faying surfaces.

In order to fully characterise corrosion in multilayer airframe structures, multiple inspections will have to be performed using techniques which can measure pillowing, material loss, the distribution of material loss by layer, and the distribution of loss on a single layer. These inspections will likely be accompanied by inspections for cracking and exfoliation corrosion around fasteners.

Data fusion can facilitate the integration of the NDI results from these multiple inspections. These techniques can also assist in the conversion of information from NDI "data" to quantitative values or probabilistic distributions that can be used by structural engineers for residual life and residual strength calculations. In terms of POD, the characteristic dimension used for establishing POD may be defined through a data fusion process rather than by a single physical quantity such as crack length or percentage thinning.

8. CONCLUSIONS

The high cost of corrosion in aircraft and safety concerns related to corrosion damage require changes to the current "find-it-fix-it" philosophy. Proactive management of corrosion will have to be supported by multimode NDI, quantified corrosion rates and corrosion damage assessment models. POD is an essential part of the characterisation of corrosion reliability.

The current subjective corrosion definitions do not permit quantitative POD values to be assigned to corrosion inspections. They are also not sufficient as inputs to DT assessments.

Corrosion metrics must be defined that allow quantitative corrosion damage assessments to be done. These metrics are fundamental to the prediction of the effects of corrosion on the static and fatigue performance of the structure. They are also fundamental to the generation of quantitative POD for corrosion. These metrics could also drive the priority for improvements in NDI for corrosion.

Quantitative characterisation of corrosion is complicated by the number of types of corrosion and by the number of different materials involved. Costs of setting up controlled POD generation programs will restrict the types of corrosion, materials and structures addressed to those which impact safety.

Data fusion techniques offer some benefits in collating the results from different inspections into a usable corrosion metric that require further study and development.

9. REFERENCES

1. "Corrosion Detection Technologies, Sector Study", Final Report, Prepared for the North American Technology and Industrial Base Organization (NATIBO), March 1998.
2. internet document: http: //www.afcpo.com/htdocs/alt/front /corrosion_bas /more_corrosion.html, US Air Force Corrosion Program Office, April 1998.
3. internet document: http://www.finishing.com/icorr/index.html, Institute of Corrosion, Bedfordshire, UK.
4. "U.S. Combat Air Power: Aging Refuelling Aircraft are Costly to Maintain and Operate". General Accounting Office, Washington, DC. National Security and International Affairs Div.; Report GAO/NSIAD-96-1608, August 1996.
5. Fabry, J. M., "Airworthiness Assurance R&D Branch 1996 Research Accomplishments", US DOT 1996.
6. Howard, M. A., Mitchell, G. O. "Non-destructive inspection for hidden corrosion in U.S. Air Force aircraft lap joints: Test and evaluation of inspection procedures", in "Structural Integrity in Aging Aircraft", American Society of Mechanical Engineers, Aerospace Division, AD **47**, 1995, pp 195-212.
7. Roach, D., "Development of a Corrosion Detection Experiment to Evaluate Conventional and Advanced NDI Techniques", 41st International SAMPE Symposium, March 24-28, 1996, pp 265-278.
8. Boeing CAG, Service Bulletin 737-53A1039.
9. Bellinger, N.C., Gould, R.W., Komorowski, J.P., "Repair Issues for Corroded Fuselage Lap Joints" – to be published.
10. Komorowski, J.P., Bellinger N.C. and Gould, R.W., "The Role of Corrosion Pillowing in NDI and in the Structural Integrity of Fuselage Joints", Fatigue in New and Aeging Aircraft – Proc. of the 19th Symposium of the International Committee on Aeronautical Fatigue Edinburgh, 16-20, June 1997, pp251-266, EMAS Publishing 1997.
11. Kinzie, R., USAF Corrosion Program Office, private communication, April 1998.
12. Wallace, W., Hoeppner, D.W., Kandachar, P.V.,"AGARD Corrosion Handbook, Volume1, Aircraft Corrosion: Causes and Case Histories", North Atlantic Treaty Organization Advisory Group For Aerospace Research and development, AGARD-AG-278, Volume 1, 1985.
13. Hoeppner, D.W., Grimes, L., Hoeppner, A., Ledesma, J., Mills, T., Shah, A., "Corrosion and Fretting as Critical Aviation Safety Issues: Case Studies, Facts, And Figures From US Aircraft Accidents and Incidents", Estimation, Enhancement and Control of Aircraft Fatigue Performance, Volume I, Editors: J.M. Grandage, G.S. Jost, Proceedings of the 18th Symposium of the International Committee on Aeronautical Fatigue 3-5 may, 1995, Melbourne, Australia, EMAS, pp.87-106.
14. Boeing CAG, 757-200 Structural Repair Manual, January 1991.
15. Curtis, D. O.,"Corrosion and fatigue in the aging fleet", Federal Aviation Administration (FAA), North West Mountain Region, 1992 .
16. Boeing CAG, "Aging Airplane – Corrosion Prevention and Control Program Model 737-100/200", July 1989.
17. Bellinger, N.C., Komorowski, J.P., "Corrosion Pillowing Stresses in Fuselage Lap Joints", AIAA Journal, Vol. 35, No. 2, pp.317-320, February 1997.
18. Goswami, T. K. Hoeppner, D. W., "Pitting Corrosion Fatigue in Structural Materials", in Structural Integrity in Aging Aircraft, ASME, AD **47**, 1995, pp 129-139.
19. Perez, R., "Corrosion/Fatigue Metrics", Proceedings of ICAF '97 Symposium on Fatigue in New and Ageing Aircraft, June 16-20, 1997, Edinburgh, Scotland.pp.215-229.
20. Doerfler M., Grandt A., Bucci R. J., Kulak M., "A Fracture Mechanics Based Approach for Quantifying Corrosion Damage", Tri-Service Conference on Corrosion, June 1994.
21. Bush, R. W., Hinkle, A. J., Bucci, R. J., Kulak M., "Prediction of the Intergranular Corrosion/Fatigue Interaction Effect in Thin-Webbed Stiffening Members", 1996 USAF Aircraft Structural Integrity Program Conference, San Antonio, TX, December 3-5, 1996. WL-TR-97-4054, pp. 337-358.
22. Goswami, T. K. Hoeppner, D. W., "Pitting Corrosion Fatigue in Structural Materials", in Structural Integrity in Aging Aircraft, ASME, AD **47**, 1995, pp 129-139.

23. Perez, R., "Corrosion/Fatigue Metrics", Proceedings of ICAF '97 Symposium on Fatigue in New and Ageing Aircraft, June 16-20, 1997, Edinburgh, Scotland.pp.215-229.
24. Doerfler M., Grandt A., Bucci R. J., Kulak M., "A Fracture Mechanics Based Approach for Quantifying Corrosion Damage", Tri-Service Conference on Corrosion, June 1994.
25. Bush, R. W., Hinkle, A. J., Bucci, R. J., Kulak M., "Prediction of the Intergranular Corrosion/Fatigue Interaction Effect in Thin-Webbed Stiffening Members", 1996 USAF Aircraft Structural Integrity Program Conference, San Antonio, TX, December 3-5, 1996. WL-TR-97-4054, pp. 337-358.
26. Forsyth, D. S., Komorowski, J. P., Gould, R. W., "The Use of Solid Film Highlighter in Automation of D Sight Image Interpretation", in "Nondestructive Evaluation of Aging Aircraft, Airports, and Aerospace Hardware II", SPIE **3397**, April 1998, pp 50-56.
27. Forsyth, D. S., Gould, R. W., Komorowski, J. P., "Correlation of Enhanced Visual Inspection Image Features with Corrosion Loss Measurements", Advances in Signal Processing for Non Destructive Evaluation of Materials, Quebec City, August 5-8 1997.

POD METHODS AND THE LINK OF AVAILABLE DATA TO FIELD PROCESSES

Ward D. Rummel
D&W Enterprises, LTD[1]
8776 W. Mountainview Lane,
Littleton, CO 80125-9406.
United States of America

1. SUMMARY

An experimental procedure is described for transferring nondestructive evaluation (NDE) procedure performance (probability of detection - POD) capabilities, that have been validated on simple specimens, to complex configurations found in field applications. Methodologies and logic are discussed. Requirements and cautions in use of the method are discussed.

2. INTRODUCTION

Increasing materials knowledge, demand for more efficient structures and systems, and demand for life-extension of aging structures and systems have prompted increasing use of damage tolerance requirements in engineering design, maintenance, rework and life-cycle management. Implementation of damage tolerance methods requires knowledge and supporting data on: (1) materials properties; (2) loads and load distribution; (3) functional operation / service cycles; (4) environment; and (5) inherent flaw sizes, locations orientations and distributions. The requirement for flaw knowledge and data is a significant addition to prior practices / art.

Flaw detection, flaw sizing, flaw location and orientation must necessarily be nondestructive in nature. The added requirement for quantification of nondestructive evaluation (NDE) measurements presents a significant challenge to the engineering community. Although some NDE capabilities have been assumed in prior designs, the assumptions were often faulty and have been shown to be inadequate by systematic measurement and quantification. Erroneous assumptions have included: (1) "no flaws assumed"; (2) an incorrect detectable flaw size assumed; (3) assumption that NDE detects "all significant flaws"; (4) detection assumed to be the "calibration" flaw size; and (5) detection assumed to be the "smallest flaw previously detected". Characterization of specific NDE procedures, NDE technicians and NDE facilities was and is required.

The metric that has been developed to quantify NDE capabilities and to provide a method of data exchange is the probability of detection (POD). Generation of a characteristic POD curve (Figure 1) requires: (1) passing a statistically significant number of representative flaws through and NDE procedure; (2) the flaw distribution must be near the expected NDE detection threshold; (3) flaws located in representative materials, geometries and surface conditions; (4) systematic control of NDE procedures; and (5) documentation of the results of application[2,3,4].

Fatigue cracks in simple test specimens are frequently used as the test artifacts. Fatigue cracks have been determined to be representative of severe detection conditions and are relatively inexpensive to produce. A large data base has been generated for NDE capabilities of relatively simple specimens[5]. Simple test specimen geometries may not be representative of the NDE challenges in a complex structure or system and methodologies for transfer of the measured capability to complex shapes are required. This paper describes such methodology and the rationale used in application. Transfer of measurements is focused on those NDE procedures which produces a quantified, scalar output such as eddy current and ultrasonic methods. The discussions are therefore intended primarily for those methods and applications.

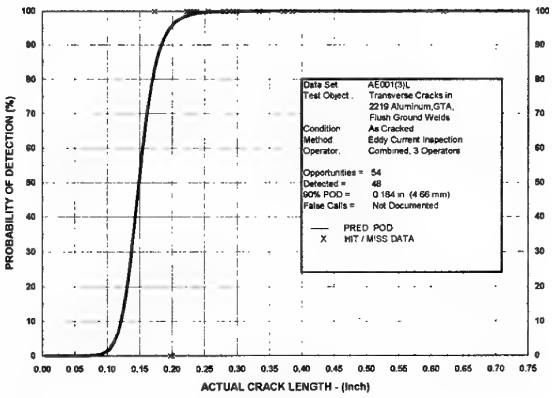


Figure 1. Typical POD curve

3. PROBABILITY OF DETECTION RATIONALE

The capability of an NDE procedure is a direct function of its signal response output from small flaws and its relationship to the background application response that is generated by unflawed areas adjacent to flaws[6]. The background response is conveniently termed the "NOISE" response and must not be confused with electronic noise that is familiar in electronic instrument analyses. When repetitive measurements of a single flaw are made by an NDE procedure, a distribution of response values from the flaw are generated that are similar to those produced in classical mechanical measurement methods. Simultaneously, a lower level signal (background) response is generated that is characteristic of the surface condition, surface texture, grain structure, stress state, etc. of the test object. This background response is termed "NOISE". A typical response from experimental measurements from a single flaw is shown in Figure 2.

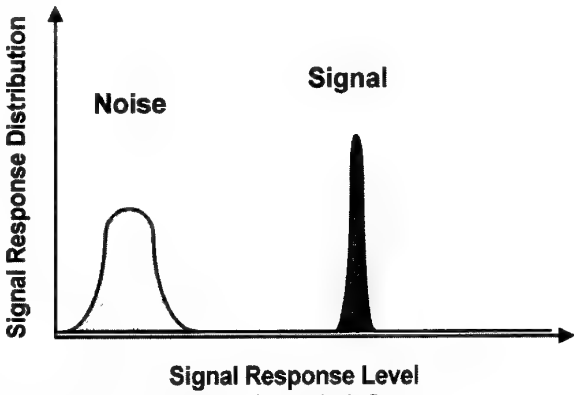


Figure 2. Repetitive response from a single flaw

Paper presented at the RTO AVT Workshop on "Airframe Inspection Reliability under Field/Depot Conditions", held in Brussels, Belgium, 13-14 May 1998, and published in RTO MP-10.

Repetitive response from multiple flaws of equal size results in broadening of the response distribution as shown in Figure 3. This broadening is the results of flaw to flaw variations and are accounted for by using multiple flaws in the generation of a typical POD curve. The spread between the upper limit of the noise and the lower limit (signal and noise) of the flaw response enables repetitive detection and discrimination / identification of flaws of that size without false calls (Type II errors). For small flaws, the signal and noise responses overlap and detection / discrimination are not attained.

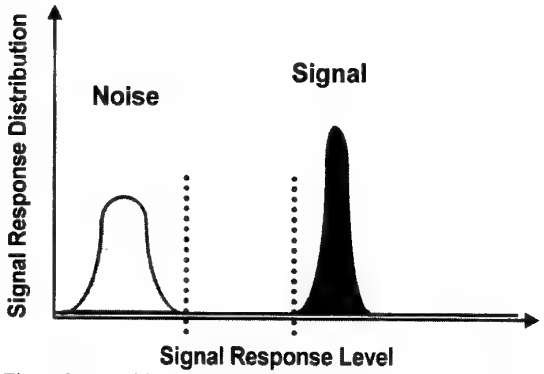


Figure 3. Repetitive response from multiple flaws of equal size

Slots generally used for purposes of set-up and "calibration" of an NDE system are more readily detected due to their higher signal response. Figure 4 show a typical response when similar measurements of a "calibration slot (artifact) are added to the response data from a crack of equal size.

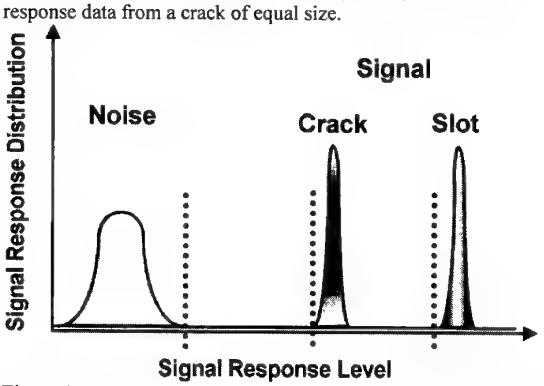


Figure 4. Comparative responses from a crack and a slot of equal size.

4. TRANSFER OF ARTIFACT RESPONSE

Most classical measurements are made with the aid of a reference "calibration" artifact or "standard". Calibration "standard" artifacts are "measured" by reference to a master standard that is traditionally retained as a as a national resource and commonality is achieved by international agreements to provide a common basis for exchange in commerce. It was therefore logical that a reference slot has evolved as a calibration artifact for most NDE measurements and physical measurement of slot size may be traceable to a national "master standard". Slots are economical to produce with available technology and are commonly specified in establishing and applying NDE procedures. Traceability of reference calibration artifacts (slots) are assumed when they are used in validated NDE procedures. Unfortunately, a single slot is often used for reference and set-up and linearity of response of the NDE procedure is assumed. Modern electronic instruments are produced with linear response and periodic validation of the response linearity is performed. Since the electronic instrument constitutes only a part of the NDE system, system capability validation for each specific application is recommended. Figure 5. illustrates a typical causal model for response to cracks and slots of varying size. For larger flaws, the response is linear. As the size of the slot / crack approaches the size of the transducer / probe element. the response function changes. It is therefore important to validate the functional response for an NDE procedure. particularly when addressing small flaws.

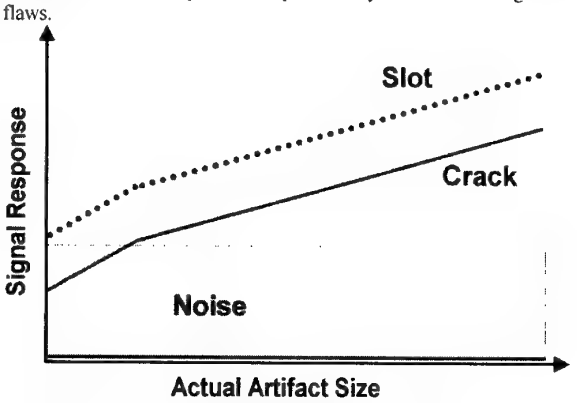


Figure 5. Typical causal response from slots and cracks

Once a relationship between responses to slots and cracks is established, a continuous function may be plotted in the form shown in Figure 5. This is the same response required in use of the **a** / **â** method used in POD generation (Response and actual crack size are plotted as logarithmic function - ln/ln).[7]

After an experimental relationship between the response of cracks of varying size and slots of equivalent varying size are established from test specimens in simple configurations, the capability of an NDE procedure for application to a complex configuration may be linked to the performance on a simple configuration using slots as the transfer artifacts. Since slots of equal physical size and shape can be economically produced in both simple and complex specimen configurations, they may be used as duplicate and traceable artifacts. A quantitative NDE response relationship may then be experimentally generated using equivalent size slots in both simple and complex specimen configurations. Care in making measurements must be exercised to link NDE performance capability (POD) based on equivalent signal response (termed "equivalent reflectivity" by some experimentalists[8,9]). Rigid control and measurement of both test specimens and data recording are required. Primary considerations include: (1) cracks used for measurements in simple specimens must be representative of the population of cracks that must be detected / measured; (2) slots used for measurements must be geometrically equivalent (size, shape, width, radius sharpness, etc.); (3) signal and noise response distributions measured must be representative of the distributions anticipated in an application; and (4) response measurements must be recorded and included in the validation data for an NDE procedure. The same slots in the complex configuration may then be incorporated into the NDE process control history by periodically determining that the response distributions for slot measurements are repeatable.

Figure 6 illustrates typical response distributions for repetitive measurements of two slots of equivalent size and the corresponding noise responses in both simple and complex specimen configurations. If the process is repeated using two slots of a different size, the same proportional relationship is obtained if the response is linear and continuous. The response relationship may then be assumed to be a constant within the bounds used in the

original crack and slot measurements. The predicted causal response for slots in the complex configuration may then be calculated over the range of crack sizes used in development of data for the simple specimen configuration. The noise data is overlaid as an upper bound limit from actual measurements made on the complex test specimen(s).

The relationship may thus be expressed as:

$$\ln\left(\frac{\text{Slot}}{\text{Response (C)}}\right)=\frac{\text{ln Slot a(C)}}{\text{ln Slot a(F)}}\left(\frac{\text{Slot}}{\text{Response (F)}}\right)$$

$$\ln\left(\frac{\text{Slot}}{\text{Response(Co}\rightarrow\text{n)}}\right)=K\ln\left(\frac{\text{Slot}}{\text{Response (Fo}\rightarrow\text{n)}}\right)$$

Figure 7 shows a calculated, continuous response for slots of varying sizes over the test range of the initial data. The corresponding noise response level is shown as an overlay as measured at the upper bound of the measured noise distribution in the complex (shape) specimen.

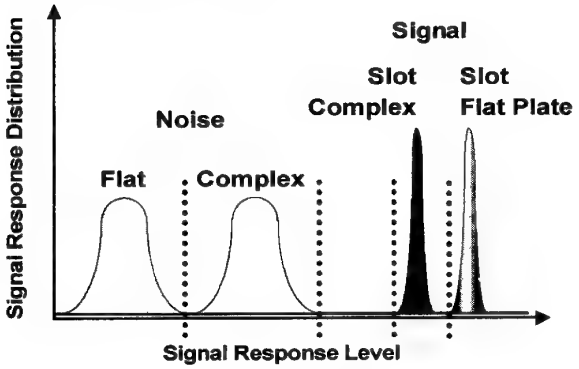


Figure 6. NDE response distributions for two equivalent size slots in a flat plate and shape (complex configuration)

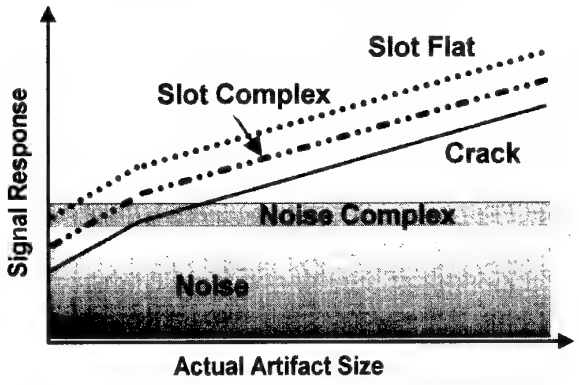


Figure 7. Calculated slot response for the complex (shape) specimen over the range or slot sizes previously quantified on flat specimens.

In like manner, response of a single crack size may be as shown in Figure 8 and a continuous crack response may be calculated from the flat plate crack data and the established slot / slot transfer constant. This relationship may be expressed as:

$$\ln\left(\frac{\text{Crack}}{\text{Response (C)}}\right)=\frac{\text{ln Slot a(C)}}{\text{ln Slot a(F)}}\left(\frac{\text{Crack}}{\text{Response (F)}}\right)$$

$$\ln\left(\frac{\text{Crack}}{\text{Response(Co}\rightarrow\text{n)}}\right)=K\ln\left(\frac{\text{Crack}}{\text{Response (Fo}\rightarrow\text{n)}}\right)$$

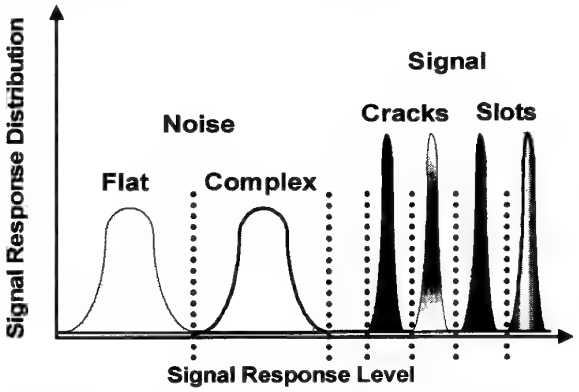


Figure 8. Calculated NDE response distributions for based on crack and slot equivalency in flat plate and complex (shape) configurations.

The response of a cracks of varying sizes in complex specimen configurations may be calculated over the same size range that was used for the flat specimens to produce a continuous response curve. The extrapolated continuous crack response is shown in Figure 9.

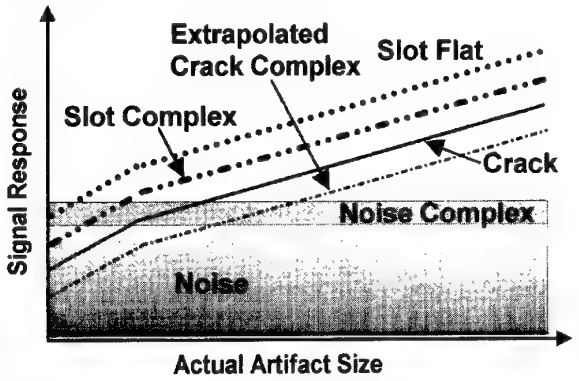


Figure 9. Extrapolated continuous crack response based on slot artifact response transfer

The probability of detection (POD) threshold crack size may be adjusted to that crack size which produces an equivalent response in the complex (shape) configuration as shown in Figure 10, A.
This method provides and equivalent POD threshold. but does not account for the change in noise, thus the false call rate would be increased. Adjustment to provide an equal false call rate and thus account for the increased noise requires setting the threshold at a point where the signal and noise margin is equal to that provided by the original flat plate data (Figure 10, B value). A new POD curves based on the extrapolated crack responses may be calculated by either the **a** / **â** or "hit / miss" methods and plotted as shown in Figure 11.

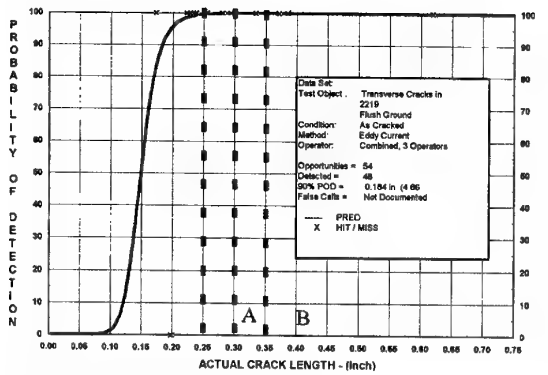


Figure 10. Adjusted POD threshold

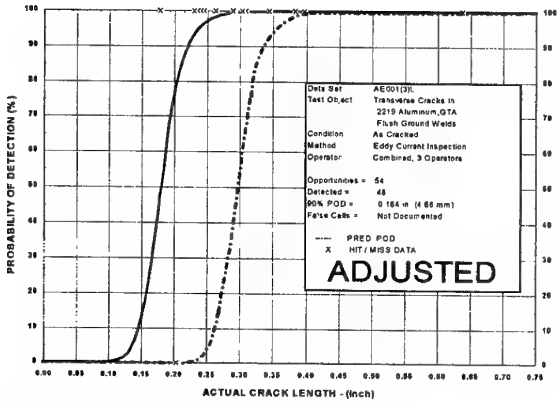


Figure 11. Recalculated and adjusted POD curve

5. CAUTIONS

Rigor is application of the method described is required and documentation of each data acquisition and calculation step is necessary for both process control and for future re-validation. Further: (1) cracks and slots must be reproducible and must be representative of the conditions under which the measurement and evaluations are to be applied; (2) physical measurements of slots and cracks must be traceable to established measurement standards; (3) all measurements must be made using the same procedure that is intended for the application; (4) crack to crack variance is not transferred and is assumed to be equal to the variance found in the flat test specimens; and (5) variances in part stress state and crack orientation are not transferred and must be addressed by the mode of application of the NDE procedure.

THE METHOD DESCRIBED DOES NOT TAKE INTO ACCOUNT ANY "HUMAN FACTORS" VARIATIONS IN APPLICATION OR EVALUATION. HUMAN FACTORS HAVE LESS IMPACT ON DISCRIMINATION LEVEL WHEN AUTOMATED ALARMS AND RECORDING ARE USED. HUMAN FACTORS FOR HAND SCANNING MUST BE ADDRESSED SEPARATELY AND INTEGRATED IN THE PROCEDURE QUALIFICATION.

6. SUMMARY

Modern design and life-cycle management require the use of damage tolerance methods and disciplines. Nondestructive detection, measurement and evaluation of both surface connected and internal anomalies is and essential part of damage tolerance methods. Nondestructive evaluation procedures must therefor be capable, reliable and quantitative in order to support damage tolerance design, acceptance and life-cycle management.

Prior to the introduction of damage tolerance methods, nondestructive evaluation procedures had not generally been rigorously characterized to establish their capability and reliability. Assumptions of capabilities were often faulty. The metric developed to quantify NDE capabilities and to provide a method of data exchange is the probability of detection (POD). POD data can be readily developed using flawed test specimens in simple configurations – often flat plates. Flawed test specimens in complex shapes and configurations are, however, difficult to obtain or may not be available or producible for new designs. A method of linking data from simple specimens to more complex applications is required.

The logic and methodologies described in this paper provide and approach to transferring nondestructive evaluation (NDE) procedure performance (probability of detection - POD) capabilities from simple test specimens to more complex applications. The methods cannot be applied in a cook book manner, but require a thorough understanding of NDE procedures, procedure characteristics, limitations and boundary conditions for application. The transfer method must therefor be considered to be a tool for use by qualified NDE engineers as a part of damage tolerance design and life-cycle management technology applications.

REFERENCES:

1. D&W Enterprises, LTD., 8776 W. Mountainview Lane, Littleton, CO 80125-9406, USA; TEL: (303) 701-1940, FAX: 791-1940 (Automatic switch)
2. A. P. Berens, "NDE Reliability Data Analysis", in Metals Handbook, 9th Edition, Vol.17, p 689,ASM International, 1989.
3. W.D. Rummel et al, "Recommended Practice for a Demonstration of Nondestructive Evaluation Reliability on Aircraft Production Part", Materials Evaluation, 40, p 922, 1982.
4. W.D. Rummel, G.L. Hardy & T.D. Cooper," Applications of NDE Reliability to Systems, in Metals Handbook, 9th Edition, Vol.17, p 674, ASM International, 1989.
5. NDE Capabilities Data Book, 3rd Edition, DB-2, 1997, available through NTIAC, (512) 263-2106.
6. Ward D. Rummel, "Considerations for Quantiative NDE and NDE Reliability Improvement, Review of Progress in Quantitative Nondestructive Evaluation, Vol., 2A, p19, 1983, Plenum Press, New York.
7. A.P. Berens, op cit.
8. R.H. Burkel, D.J. Sturges, R.S. Gilmore and W.T. Tucker, "Effective Reflectivity: POD Methodology for Ultrasonic Inspection, Paper presented to the 1995 Fall Conference of the American Society for Nondestructive Testing, Dallas, Texas.
9. Olav Forli, et al, "Guidelines for replacing NDE techniques with one another, NT Report 300, NORTEST, P.O. Box 116, FIN-02151 ESPOO, Finland, 1995.

AN EVALUATION OF PROBABILITY OF DETECTION STATISTICS

D. S. Forsyth
A. Fahr
Institute for Aerospace Research
National Research Council
Building M14, Montreal Road, Ottawa ON Canada K1A 0R6
email david.forsyth@nrc.ca

SUMMARY

Statistics and methodologies used to develop probability of detection (POD) information are examined with examples from data sets obtained by inspecting service-retired engine components.

The effects of using different statistical methods to analyze POD data are demonstrated. As the study of nondestructive inspection (NDI) reliability has matured, different methods for the design of reliability experiments and analysis of resulting data have been proposed. The application of different methods to the same POD data set is evaluated. Log normal and log odds (also called logistic) models are shown to yield very similar results when parameters are estimated using maximum likelihood estimation (MLE) techniques. The use of range-interval techniques for parameter estimation yields poor results.

The use of repeated inspections for improving reliability is discussed. Finally, the importance of using representative flaws for POD studies is demonstrated.

LIST OF SYMBOLS

IAR - Institute for Aerospace Research
MLE – maximum likelihood estimation
POD – Probability of Detection
NDI – nondestructive inspection

1. INTRODUCTION

The NDI group at the Institute for Aerospace Research (IAR) has carried out a number of studies of the reliability and sensitivity of NDI techniques applied to aerospace components (e.g. Refs. 1, 2, 3). In reliability studies, inspection data is transformed to a relationship between the probability of detection of flaws and a characteristic size of the flaws. The discrete sample data set can be used to estimate the global population statistics, and log odds or log normal curves have been used to model this relationship. Different methods have been suggested to estimate the parameters which describe either the log odds or log normal models. In this work, some of these options are evaluated on data obtained by inspecting service-retired components.

2. NDI RELIABILITY EXPERIMENTS

The formal study of the reliability of NDI is relatively new, with some of the first applications being in the NASA space shuttle program in the early 1970's (e.g. Ref. 4).

The design of experiments to determine the POD of an NDI system has been thoroughly addressed in References 5 and 6. Three categories of experiments have been used to evaluate the reliability of NDI.

2.1 Category 1: Demonstration at One Flaw Size

Historically, some experiments have been performed to demonstrate the ability of an NDI system by using multiple test specimens with the same size flaw. Based on statistical sampling theory, 29 successes in 29 trials at one flaw size gives a 90% confidence that this flaw size will be found every time. This method provides much less information about the inspection system than using a range of flaw sizes for test specimens. It is used mainly to satisfy regulatory concerns.

2.2 Category 2: Estimation of POD Using Single Inspections

To qualify the application of an NDI system to a particular problem, an experiment of this type must be performed. The results of a properly designed experiment will allow the generation of a valid POD curve.

One method of performing this type of experiment is to analyze NDI data from fleet inspections (Ref. 7). The results of the inspection of a single subject, for example one fastener hole, are recorded until a flaw is detected and repair or replacement takes place. Crack growth data is then used to estimate the size of the flaw at the times of previous inspections that may have missed this flaw. If the same inspection has been performed at intervals of time on the same subject, there may be a set of NDI results for different crack sizes. Given a large enough sample population, it may be possible to estimate a statistically valid POD.

The more common method of performing NDI experiments involves using simulated flaws or components with real service-induced flaws (Ref. 5,6). Again, given a distribution of flaw sizes and a sufficient number of flaws, a valid POD can be found.

2.3 Category 3: Estimation of POD Using Multiple Inspections

It is possible that by using multiple inspections a greater POD may result than from using one technique alone. An experiment of this type is essentially the same as performing multiple Category 2 experiments, that is, using more than one NDI system on the same subject. It should be noted that the use of multiple inspections has been shown in some cases to provide no benefit to the POD.

2.4 NDI Reliability experiments at IAR

Data reported in this paper are taken from a recent study of the reliability of NDI techniques applied to the detection of low cycle fatigue (LCF) cracks in engine compressor disks (Ref. 3). This study was initiated to optimize an automated eddy current system, ARIES, built by Tektrend International under a contract from IAR. Other inspections were carried out at IAR and at various commercial NDI operators.

Paper presented at the RTO AVT Workshop on "Airframe Inspection Reliability under Field/Depot Conditions", held in Brussels, Belgium, 13-14 May 1998, and published in RTO MP-10.

This was a Category 2 experiment based on the above definitions. However, because multiple independent inspections were performed on the same specimens, Category 3 experiments can be derived from the results by combining selected results of individual inspections.

The NDI results for this particular experiment were only reported in terms of "hit" or "miss", that is, no estimation of crack size was reported. Although most of the techniques used could have provided such an estimate, this is not commonly done at the depot level for these components.

Many different inspection techniques were used in this trial, and subsets of these were carried out at different organizations with different inspectors.

2.5 Specimens for NDI Reliability Experiments

There are many ways to collect the data required to determine the reliability of an NDI technique in a particular application. These range from using field service data, using service retired components as in this study, to using artificially created flaws in real or simulated components, to using generic test blocks. Unfortunately, the best reliability data will be the most expensive, requiring a significant number of flawed and flaw free components. This also requires that components are used in service before accurate determination of NDI capability. Because of these limitations, NDI reliability is often estimated by using artificially generated flaws or even computer models.

As part of this study, EDM notches were created in virgin holes drilled in the test components, and fatigue cracks were generated in material removed from the test components. The response of eddy current instruments to these different types of flaws was evaluated.

3. POD MODELS

3.1 Log Odds

Berens and Hovey (Ref. 8) examined various methods of modeling NDI data to determine POD curves. They concluded that the log odds distribution was the most consistent distribution for determining a POD curve as a function of crack length a_i. The functional form of the log odds distribution is as follows:

$$P_i = \frac{\exp(\alpha + \beta \ln(a_i))}{1 + \exp(\alpha + \beta \ln(a_i))} \qquad (1)$$

where P_i is the probability of detection for crack i, a_i is the length of crack i, α and β are constant parameters which define the curve.

Berens and Hovey (Ref. 8) presented two approaches for estimating these constants, the Range Interval Method (RIM) which is also known as Regression Analysis and the method of Maximum Likelihood Estimators (MLE). Reference 5 also suggest the calculation of confidence bounds on the log odds POD curve by assuming that the estimates of the mean POD curve will be normally distributed about the true POD curve, for large sample sizes.

3.1.1 Log Odds: Using the Range Interval Method

The Range Interval Method (RIM) has been used to estimate the parameters α and β required to define a log odds curve (see equation (1)). It is assumed that the variability of POD within a small crack size range or interval is small and the detection within that range follows a binomial distribution (Ref. 8). To implement the range interval method, the crack data is divided into t intervals of equal length. The probability of detection is calculated for each interval as being the ratio of cracks detected to the total number of cracks in that interval. This gives t data points.

The t data pairs of POD and crack length are transformed into a linear domain and a linear regression is performed on the data pairs in order to obtain the intercept and slope parameters, α and β, of the log odds function (equation (1)). The reverse transformation gives the POD curve.

The data points are transformed into a domain where the POD relationship is linear, using the following transformations on the log odds distribution function:

$$Y_i = \ln\left(\frac{P_i}{1 - P_i}\right), \; X_i = \ln(a_i) \qquad (2)$$

where P_i is the proportion of cracks detected and a_i is the crack length in the interval i.

The result of the transformation on equation 2 is a set of points which are fitted with the line:

$$Y = \alpha + \beta X \qquad (3)$$

These parameters α and β can be substituted into equation 1 and used to calculate a POD curve for a range of crack lengths. It should be noted that if the estimated POD value for an interval is 0 or 1, this transformation is undefined. For a POD of 0, a value of 1/(t+1) is used, for a POD of 1, a value of t/(t+1) is used for the transform. This approximation overestimates the POD at small crack sizes where the POD in an interval is zero, and underestimates the POD at large crack sizes where the POD in an interval is 1.

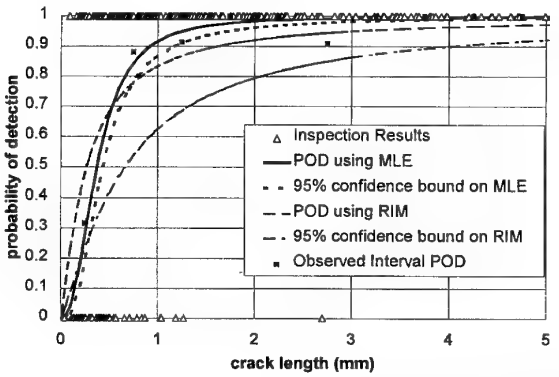


Figure 1. A comparison of RIM and MLE methods of fitting log odds curves to inspection data.

Figure 1 shows a comparison of the RIM and MLE methods of fitting the log odds curve to actual POD data. Also included are the values of POD for the intervals used to fit the RIM curve. Based on previous work (Ref. 9), use of the RIM method is not recommended. Results of the RIM curve-fitting are very sensitive to assumptions made in the execution of the curve-fitting algorithm.

3.1.2 Log Odds: Using Maximum Likelihood Estimators

The MLE technique is used in this application to find estimates of the parameters α and β from equation (1) that maximize the

probability of obtaining the observed data. The likelihood L for a single observation is:

$$L(P_i; a_i, x_i) = P_i^{x_i} \bullet (1 - P_i)^{1 - x_i} \qquad (4)$$

where P_i is the probability of detection for crack i, a_i is the length of crack i, and x_i is the inspection outcome, 0 for a miss and 1 for a hit.

The likelihood of a series of independent inspections is the product of the individual observations:

$$L(P; a, x) = \left[\prod_{i=1}^{h} P_i\right]\left[\prod_{j=1}^{n-h} (1 - P_j)\right] \qquad (5)$$

By taking the logarithm of equation (5), the series of products becomes a series of sums, equation (6). The logarithm is a monotonic function, so the maximum of the log likelihood for α and β is the same as the maximum of the likelihood.

$$\ln L(P; a, x) = \sum_{i=1}^{h} \ln P_i + \sum_{j=1}^{n-h} \ln P_j \qquad (6)$$

Equation (6) is differentiated with respect to α and β, derivatives set to zero, and the resulting simultaneous equations are solved. This gives the estimates of α and β that maximize the likelihood.

Figure 2 shows a log odds and log normal curve fit to the results of an ultrasonic inspection of bolt holes in turbine disks. The log odds fit was performed using the MLE method.

3.2 Log Normal

The cumulative log normal distribution is suggested by Petrin et al. (Ref. 5) for modeling POD data. The cumulative log normal distribution is expressed as:

$$P_i = 1 - Q(z_i)$$

$$\text{for } z_i = \frac{\ln(a_i) - \mu}{\sigma} \qquad (7)$$

where Q(z) is the standard normal survivor function, z_i is the standard normal variate, and μ and σ are the location and scale parameters of the POD curve.

The MLE method can be used to find the values for the location and scale parameters. The same likelihood function, equation (6), applies to both the log odds and log normal fit. Equation (6) is differentiated with respect to μ and σ, derivatives set to zero, and the resulting simultaneous equations are solved. This gives the estimates of μ and σ that maximize the likelihood.

The method of determining the confidence bound on the log normal POD curve is derived by Cheng and Iles (Ref. 10).

Figure 2 shows a log normal and a log odds curve fit to the results of an ultrasonic inspection of bolt holes in turbine disks. The log normal fit was performed using the MLE method.

3.3 Comparison of Log Odds and Log Normal Curve Fits

In general, the log odds and log normal curve fits are very similar for the data obtained in this set of trials. At lower values of POD, less than about 0.5, the log normal curve fit produces a smaller crack length. But at higher values of POD, such as the 90% value often used, the log normal is usually slightly more conservative, meaning for a given POD the corresponding crack size is larger for the log normal curve fit than the log odds. Table 1 shows the difference between the log odds and log normal fits at the 50% POD and 90% POD for a few different inspections. Figure 2 shows an example of the entire POD curve for an ultrasonic inspection, the same ultrasonic inspection referred to in Table 1. Details of the inspection procedures can be found in Reference 3.

The difference between these two curve fits is not significant over much of the range of POD in the data reported herein. However, it should be noted that the differences are greatest at the extremes of the curve, which is important because of the use of the "90/95" point in many design criteria. These differences are also larger as the slope of the POD curve moves away from vertical at the POD 0.5 point.

Table 1. A comparison of the log odds and log normal curve fits.

	crack length at 50% POD (mm)		crack length at 90% POD (mm)	
Technique	log odds	log normal	log odds	log normal
ECI – A, P	0.39	0.38	0.75	0.79
ECI – M	0.39	0.39	0.74	0.77
UTI	1.13	1.13	1.54	1.67
LPI	2.25	2.29	3.45	3.94

Key:
ECI - A,P: automated eddy current system, automated interpretation
ECI – M: manually operated eddy current, manual interpretation
UTI: ultrasonic inspection
LPI: liquid penetrant inspection

The choice of which curve fit to use is still a matter of debate. To facilitate the exchange of POD data, it is important that organizations carefully reference the statistics used in generating this kind of information. An accepted method for estimating the goodness of fit of these curves to the actual inspection data could suggest a preferred curve fit.

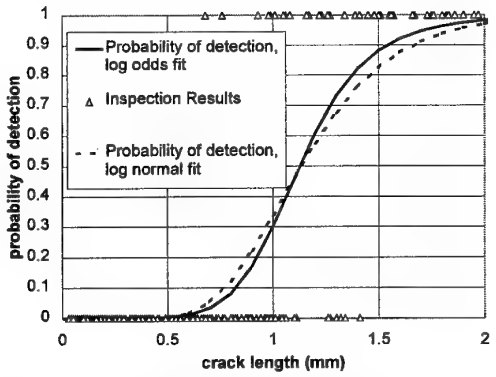


Figure 2. Log odds and log normal curve fits to results of an ultrasonic inspection of compressor disk bolt holes.

4. MULTIPLE INSPECTIONS

The use of multiple inspections to improve POD has been suggested. A simple analysis that assumes complete independence shows a large benefit from multiple inspections. A more reasonable assumption is that there is some dependence, which can be measured (Ref. 11) by making multiple inspections. Data sets of multiple inspections of the bolt holes in these trials were formed after all trials were complete. As stated in section 2.4, different NDI techniques were employed at different organizations during the course of this study. Therefore sets of multiple inspections can be made where different inspectors used different techniques at participating organizations, which should maximize the independence between inspections.

The multiple inspection data is generated by combining individual inspection results using the logical OR, that is, if any inspection finds the crack, the combined inspection is considered to have found the crack.

The first example combines the results of the automated eddy current system (ECI-A,P) and the ultrasonic inspection (UTI) from Table 1. No POD curves are shown for this case, because the ECI-A,P inspection found all the cracks that were found by the UTI inspection. Thus there was no improvement by performing both inspections.

A second example was created by combining both eddy current inspections shown in Table 1. The ECI-A,P inspection found 19 cracks that were not found by the ECI-M inspection. The ECI-M inspection in turn found 18 cracks not found by the ECI-A,P inspection. The inspection result and POD curve fit are shown in Figure 3. The log odds and log normal POD fits to this inspection were very similar, for clarity only the log odds is shown

The combined eddy current inspections were more sensitive than the individual inspections, however the false call rate was higher. Table 2 shows a comparison of the crack lengths at a POD of 0.9, and at the 95% confidence level for POD of 0.9 (the "90/95" length), for both log odds and log normal fits.

The use of multiple inspections to improve NDI reliability relies on there being some independence between the inspections. It was shown that a sensitive eddy current test was not improved by the addition of a less-sensitive ultrasonic test. However, when two eddy current tests of similar sensitivity were combined, there was a slight improvement in the POD.

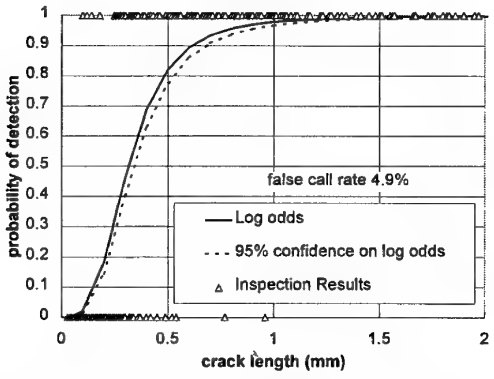


Figure 3. The results of a multiple inspection: an automated and a manual eddy current inspection.

This independence between inspections is likely due to the variability in inspection procedures. A highly manual technique is likely to have more random variability than an automated one. Thus the replacement of manual techniques with automated techniques should reduce the need for multiple inspections by increasing the reliability of single inspections. Automated eddy current techniques have been shown to be more sensitive and reliable than manual techniques in laboratory settings (Ref. 3), and this reliability is more likely to be maintained in the transfer to a hangar environment.

Table 2 . A comparison of individual eddy current results with combined results.

	crack length at 90% POD (mm)		crack length at 90% POD at a 95% confidence (mm)	
Technique	log odds	log normal	log odds	log normal
ECI – A, P	0.75	0.79	0.80	0.87
ECI – M	0.74	0.77	0.80	0.84
combined	0.62	0.64	0.68	0.71

Key:
ECI - A,P: automated eddy current system, automated interpretation
ECI – M: manually operated eddy current, manual interpretation
combined: results of ECI – A,P and ECI – M combined using the logical OR operation

5. RESPONSE OF EDDY CURRENT INSPECTIONS TO DIFFERENT FLAW TYPES

As previously mentioned, simulated flaws of two types were generated in specimens cut from the disks under study, for comparison with actual LCF cracks that developed in service. These were EDM notches, and fatigue cracks grown from starter EDM notches in a laboratory. The fatigue cracks were started in a hole smaller than the actual bolt holes, which were drilled out to the same size as the bolt holes after crack growth was complete. This eliminated the depth of the EDM starter notch from study.

Eddy current inspections were carried out on ten EDM notches of various depth, ten laboratory grown fatigue cracks of various depth, and a subset of the service-induced LCF cracks. The relationship of the maximum amplitude of the eddy current response to the crack face area is shown in Figure 4 for the three flaw types.

This comparison was performed using an Elotest B1 instrument with a 4.7 mm diameter probe manufactured by NDT Instruments for this particular application. This is the equipment used in IAR's ARIES eddy current inspection system referred to previously.

The in-service flaws were all highly oxidized in the engine operating environment, which likely resulted in a higher impedance across the crack face than the laboratory grown fatigue cracks. The EDM notches had a significant air gap which was not seen in either the in-service or laboratory grown fatigue cracks. These factors made for significant differences in the response of an eddy current system to the same size flaw.

For different inspection techniques, the oxidation of the crack face might have little or no effect. However, the tightness of the fatigue cracks in comparison to the EDM notches would very likely result in different responses for ultrasonic or penetrant methods applied to EDM notches and fatigue cracks.

If artificially generated specimens and flaws are to be used to develop reliability data for NDI techniques, the differences between the in-service flaws and the artificial flaws must be understood and accounted for.

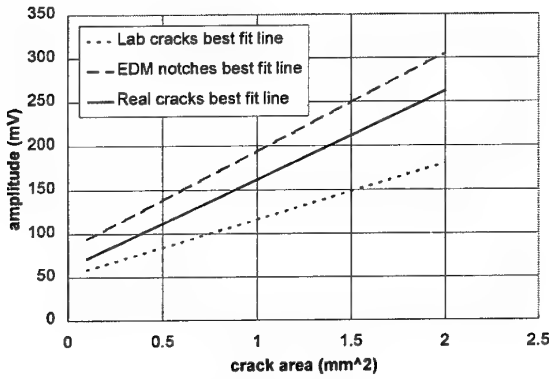


Figure 4. A comparison of the eddy current signal response to different flaw types.

6. CONCLUSIONS

Many options are available for NDI reliability experiments, with the most realistic data being the most difficult to obtain. The log odds and log normal models of the relationship between POD and crack size produce similar results, but they are not the same.

The use of multiple inspections has been suggested to improve POD. In some cases, it does slightly improve the POD. However, the use of automation in NDI should reduce the random component of NDI performance, which would also reduce the benefit of multiple inspections.

Careful attention must be paid to the type of specimens used for NDI reliability trials, as demonstrated by the differences in eddy current response to flaws of the same size shown in Figure 4.

7. ACKNOWLEDGEMENTS

This study was carried out with IAR support under Structures, Materials, and Propulsion Laboratory sub-program 318, project JHU-00; with financial support from the Canadian Department of National Defence, DAS Eng 6-2.

8. REFERENCES

1. Bellinger, N. C. et al., "The Reliability and Sensitivity of NDI Techniques in Detecting LCF Cracks in Fastener Bolt Holes of Compressor Discs", NRC NAE LTR-ST-1651, February 1988.
2. Fahr, A. et al., "POD Assessment of NDI Procedures Using a Round Robin Test", AGARD-R-809, January 1995.
3. Forsyth, D. S. and Fahr, A., "The Sensitivity and Reliability of NDI Techniques for Gas Turbine Component Inspection and Life Prediction", NRC IAR LTR-ST-2055, July 1996.
4. Rummel, W. D., Todd, P. H. Jr., Frecska, S. A., and Rathke, R. A., "The Detection of Fatigue Cracks by Nondestructive Testing Methods", NASA CR-2369, February 1974.
5. Petrin, C., Annis, C., and Vukelich, S. I., "A Recommended Methodology for Quantifying NDE/NDI Based on Aircraft Engine Experience", AGARD-LS-190, April 1993.
6. Spencer, F., Borgonovi, G., Roach, D., Schurman, D., and Smith , R., "Reliability Assessment at Airline Inspection Facilities", Vol. I : A Generic Protocol for Inspection Reliability Experiment, DOT/FAA/CT-92/12, I, March 1993.
7. Simpson, D. L., "Development of Non-destructive Inspection Probability of Detection Curves Using Field Data, NRC NAE LTR-ST-1285, 1981.
8. Berens, A. P. and Hovey, P. W., "Characterization of NDE Reliability", *in* Review of Progress in Quantitative NDE, *1*, New York, Plenum Press, 1982.
9. Fahr, A., Forsyth, D. S., and Bullock, M., "A Comparison of Probability of Detection (POD) Data Determined Using Different Statistical Methods", NRC IAR LTR-ST-1947, December 1993.
10. Cheng R.C.H., and Iles, T.C., "One sided Confidence Bands for Cumulative Distribution Functions", Technometrics, *30*, 2, 1988, pp 155-159.
11. Erland, K., "Quantifying the Benefit of Redundant Flourescant Penetrant Inspection", *in* Review of Progress in Quantitative NDE, *8B*, New York, Plenum Press, 1988, pp 2221-2228.

Identifying Sources of Variation for Reliability Analysis of Field Inspections

Floyd W. Spencer
Sandia National Laboratories, Building 957
P.O. Box 5800, Mail Stop 0829
Albuquerque, New Mexico 87185-0829
USA

1. SUMMARY

A consistent finding across many reliability programs is that inspector-to-inspector differences constitute a major source for variation. This inspector-to-inspector variation, however, can be due to many factors. Understanding individual components of variation for the observed inspector-to-inspector variation is essential, if that variation is to be reduced. Two categories of factor are equipment related and decision related. Equipment related factors include not only the settings of the NDI equipment (gates, gains, etc), but also the relationship of the inspection equipment to the material to be inspected (alignments, coupling, etc.). Decision related factors are those things that influence the inspector's call based on a signal from the inspection.

A combination of laboratory and field experiments is useful for studying the impact on inspection reliability that is achievable in the field. Although not directly reflecting all conditions that can influence an inspection, the laboratory environment usually gives the researcher an opportunity to study quantitatively the effects of equipment related variables on signal responses. This should be done using appropriate statistical design of experiments, such as full factorials and fractional factorial designs.

Inspection results taken in the field will usually consist of call – no call data and gives the researcher a chance to see the total effect of many factors on the inspection process as implemented. However, when possible, data concerning the factors studied in the laboratory should be gathered during the field inspections. This will enable a direct comparison to the laboratory results and provide the chance to assess whether field inspectors are inducing more variation than expected in setups and other equipment related variables. This, in turn, will allow a more direct assessment of the decision processes being used in the field environment.

With respect to POD curves, the hit – miss or call – no call data that is gathered in field conditions should be analyzed allowing for processes that will result in hits and misses that may arise for reasons other than directly related to a crack length. The most direct way to do this is to generalize the usual 2 parameter POD curves to four parameters. The additional parameters set levels of hits and misses that are independent of crack length and results in POD curves that start at values greater than zero and approach an upper limit other than one.

The extension of traditional models of POD to include two additional parameters will require special concerns in the design of experiment for field inspections. This is especially true for setting up appropriate crack distributions to be used in these experiments.

2. INTRODUCTION

It has been recognized that nondestructive inspection (NDI) techniques and instruments that have proven themselves in the laboratory do not always perform as well under field conditions. In this paper we explore combinations of formal laboratory and field experimentation to characterize NDI processes as they may be implemented in field conditions.

We also discuss appropriate modeling for probability of detection (POD) curves as applied to data gathered under field conditions. A case is made for expanding the more traditional two-parameter models to models using either three or four parameters. We use NDI data gathered from various airframe inspection programs to illustrate the points.

3. DESIGN OF EXPERIMENTS FOR TECHNIQUE CHARACTERIZATION AND FIELD POD

Reliability programs that gather data across many users consistently find inspector-to-inspector variation to be substantial [1]. The term, *human factors*, is often used to encompass this variation, but in such a way as to imply that the ultimate cause of the variation is a mixture of psychological and physical conditions that are specific to the inspector and the inspector-environment interaction at the time of inspection. However, a true understanding of inspector-to-inspector differences in reliability is more likely if the parameters of the inspection that can cause variability are quantitatively understood.

Paper presented at the RTO AVT Workshop on "Airframe Inspection Reliability under Field/Depot Conditions", held in Brussels, Belgium, 13-14 May 1998, and published in RTO MP-10.

Procedural factors that can impact the outcome of an inspection are best studied in a laboratory environment where control of factors can be maintained according to Design of Experiment (DOE) principles. In the laboratory environment one usually can gather signal response data as a function of input variables. This is opposed to field data that is usually calls and no-calls or hits and misses when combined with knowledge of flaw locations.

When possible, data specific to individual setups used in field inspections should be gathered for direct comparison to the values used in the laboratory characterization. The concepts are discussed more fully in the following and are illustrated with recently completed experimental programs, the details of which are presented in references [1,2].

One of the programs [2] used to illustrate some of the ideas presented here was a program to develop and validate an ultrasonic inspection system for locating 2nd layer cracks in the lower inner-wing spanwise splice-joints of C-141 aircraft. The cracks occur at fastener sites.

3.1 Laboratory Experiments

The intent of the laboratory validation is to characterize the impact of procedural variables not only on detection, but also on the quality of the signal. The variables to be included in the experiment should include those that are expected to vary with each setup and implementation of an inspection procedure. Once those factors are identified, the amount of expected variation in those factors needs to be established. Then, well known statistical design of experiment concepts, such as factorial and fractional factorial plans, can be used to characterize the impact of those variables on the signal that will be used to make a call.

In the example program [2] it was determined that there were 5 variables that would consistently differ from one inspection to the next. These variables were time base delay, depth velocity, receiver gain, skew of scanner travel in relation to the fastener sites, and the applied probe pressure.

Levels for each of the factors were determined by analyzing the setup and calibration procedure steps. For each variable, high and low levels were determined as setting the range of values that could be expected when procedures were followed. A one-half fraction factorial experiment of 16 runs (2^{5-1}) was augmented with a run at nominal levels for the variables to define the input variable levels that would be used in the laboratory experimental program. The runs were then blocked in two groups of 8, with each block being carried out on a different set of specimens. (The nominal run was performed on both sets of specimens.) The result was that each run encompassed approximately 180 fastener sites. Signals were recorded for all the inspections so that the effect of the input variables could be analyzed with respect to various signal characteristics.

The results of a laboratory experiment can be used in several ways. If performed early in an NDI development program, the results can be used to specify acceptable levels and controls on procedural variables, including setups and calibrations. This will help to assure that variations in inspection results are controlled to an acceptable level. For existing procedures, the results of such a laboratory experiment can establish bounds on observed field variation that can be attributed to specific factors and not just attributed to *inspector-to-inspector* or *human factor* effects.

3.2 Field Experiments

We use the term *field experiments* to apply to situations in which data are gathered on inspections in an environment and under conditions closely related to actual inspection conditions. Test specimens that are used as inspection articles may be specially fabricated to achieve some control over the distribution of flaws, but the conditions under which the test articles are inspected should be as close as possible to conditions that are expected for the routine implementation of the inspection technique.

In order to have inspection conditions that were realistic, the program to assess the reliability of high frequency eddy current inspection in airline maintenance facilities [1] designed special frames to hold the test specimens in a manner to simulate the side of an aircraft. The experiment was then located in various facilities in the way an airplane would come in for an inspection. The net result was that the data were gathered with the individual inspectors following the same procedures and operating in the same environment that would result from an airplane being brought in for an inspection.

Similarly, in the C-141 program [2], the test specimens and support structure were designed in such a manner that required the inspector to operate the inspection just as he would have to on the bottom side of the aircraft wing. Thus, the inspectors were required to perform all the procedural steps of placing and attaching the scanner to the underside of a wing.

The various factors that were identified for laboratory experimentation will not be controlled in the field portion. They will be allowed to occur at levels established by the inspectors. This does not mean that

the field experiment is void of statistical design principles. Inspector specific traits, such as experience and training should be considered as possible variables.

For the C-141 program, the inspection technique being characterized was newly developed. Therefore, there were no inspectors with experience with the specific application. However, inspectors that would be called upon to perform the inspections could be characterized according to experience levels with the automated ultrasonic imaging technique that was being deployed. Three groups were identified, based upon the amount of training that was deemed necessary. These were expert, intermediate, and novice groups that would receive 1 day, 1 week, and 2 weeks of training respectively. By insuring that these groups were represented in the field experiments, there would be data on the efficacy of the various training programs.

3.2.1 Correlating field results with laboratory
Although the procedural type variables that are studied in the laboratory will not be controlled in the field experiments they can still be recorded and compared to the levels used in the laboratory. The field data provides a check for the adequacy of the range of inputs used in the laboratory.

For the C-141 program the time base delays and gains used by the inspectors varied more than the range used in the laboratory characterization. The laboratory range was based on an analysis of what should be expected from strict adherence to the procedures. Identifying the reason for the added field variation is a concrete step toward improving the inspection system.

3.2.2 Procedure Implementation
The field inspections should include all major procedural steps that could possibly influence the reliability of an inspection. This includes calibration and setup of equipment, as well as the handling of equipment during an inspection. In the C-141 program this meant that the inspectors had to attach, with suction cups, a two-axis scanner to the underside of a wing surface. Proper scanner attachment and the appropriate definition of the inspection area in the computer were essential steps.

The field experiments provide the opportunity to observe discrete events that could impact reliability and that may not be uncommon. Examples in the C-141 program include the reverse mounting of a transducer following calibration and scan misalignment that effectively removed the last fastener in the scan from being inspected.

If possible, the field experiments should address both the skill and mechanical aspects of an inspection as well as the decision process once a signal is obtained. This separation of the inspection tasks was easily accomplished in the C-141 program because all signal images obtained by the inspectors were saved. However, the program went an additional step to separate the data acquisition process from the decision process by asking some of the inspectors to make calls on a stored image set. The process to access those images was similar to the process that they were taught for an actual inspection. The difference was that they did not have to perform the actual inspection. In the C-141 case there was substantial variation in the calls made on this common data set.

The variation of calls made on the common data set was nearly as great as the calls made from the individual inspections of the test specimens. Clearly, assuring that inspectors applied a more uniform decision process would increase the reliability of the inspection. This issue could be addressed through improvements in training as well as the possible development of computer aided decision tools.

4. POD MODELS TO REFLECT FIELD DATA

In the previous section we discussed design of experiment philosophies for integrating laboratory and field data into reliability assessments. Results of such experiments are usually summarized by probability of detection curves, where a probability is established as a function of a flaw characteristic. For the purposes of the following discussion we use crack length as the flaw characteristic, although in practice some other variable may be more appropriate.

Data from laboratory experiments are more likely to be able to be gathered as variables, as opposed to binary. Data from field experiments are likely to be hit/miss or call/no-call data. The usual analysis of either form of data is addressed in the literature [3]. Further discussions and software have been made available in US Air Force sponsored programs [4,5].

The cited references refer to the variable response analysis as an a-hat versus a analysis. In this context, a-hat is a single inspection variable that is treated as being directly related to the crack length, a. We will not pursue this form of analysis other than to note that extensions to multidimensional data can be made [6].

4.1 Probability of Detection Curves for Hit/Miss Data

The usual probability of detection curves are assumed to be monotonic and to go from 0 to 1 as a function of the crack length. Implicit in this modeling is the assumption that if a crack is small enough it will have probability of 0 of being detected and if it is large

enough it will have a probability of 1 of being detected. The curves used to model POD are usually two parameter functions. The two most common are derived from using log-logistic and lognormal probability distribution functions. We will not repeat the mathematical forms here, but note that the two parameters control the location of the POD curve (that is, where the 50% detection rate is) and the scale (how fast the POD changes as a function of the crack length).

Probability of detection curves are empirically estimated from the hits and misses made on a set of test specimens with a range of crack lengths. If an inspection technique called everything as flawed, then an estimate of POD would be 1 regardless of flaw size. Such a procedure would also be calling non-flaws as being flawed and therefore would be yielding a high false call rate.

Some have integrated false calls into the equation by modeling what is referred to as the probability of an indication [7]. This model is given by,

$$POI(a) = p + (1- p)\cdot POD(a), \qquad (1)$$

where a is crack length and p is the probability of a false indication or false call rate. This POI model starts at p for small cracks and goes to one for large cracks.

In reference [1] it was pointed out that some of the misses for large cracks were for reasons other than a lack of an appropriate signal by the NDI technique. This lead to modeling the probability of detection as,

$$POD(a)= (1-p_m)\cdot F(a;\mu,\sigma), \qquad (2)$$

where p_m is a probability of miss independent of crack length, a is the crack length, and F is one of the usual two-parameter distribution functions (log-logistic or lognormal) that is fit to hit/miss data.

Equations (1) and (2) suggest a more general 4-parameter model given by,

$$POD(a) = p_h + (1- (p_m+ p_h))\cdot F(a;\mu,\sigma), \qquad (3)$$

where p_h is the "false call" rate, p_m is the probability of a missed call independent of crack length, and the $F(\cdot;\mu,\sigma)$ function is a distribution function modeled with two parameters. Such a POD function has a lower asymptote of p_h and an upper asymptote of $1- p_m$. All parameters can be estimated by the maximum likelihood method.

Except for the substitutions of POD for POI and F for POD, equation (3) looks like a generalization of equation (1). There is however, a major philosophical difference in the two forms that should be discussed. Users of the model given in equation (1) will re-express the POD in terms of the POI and the parameter p. The implication is that the "true" detection process is simply overlaid with another random process that introduces a nuisance parameter p. Once the nuisance parameter has been estimated it can be removed to reveal an actual POD. However, there is no *a priori* reason to believe that the constant detection rate, p, is not an inherent part of the process being modeled. If it is part of the process then modifications made to alter it could very well be altering the rest of the equation as well.

We consider equation (3) as providing a mathematical framework that has desirable properties for modeling a probability of detection. However, we draw no distinction between an "indication" and a "detection" for flaws. If a flaw was called in an inspection then it was detected. Whether the function, F( ), has an inherent meaning is a question whose answer is dependent upon the application and conditions under study.

Figure 1 shows four idealized density curves for the distribution of an NDI signal from an inspection procedure for finding cracks in a specific application. Starting from the left, the first three curves are all bimodal. The second mode for each is the small rise on the right. The curves represent signal distribution for noise, a small crack, a moderate crack, and a large crack. The vertical lines represent potential thresholds that are used to make calls. The term, *noise,* is used in this context as any signal resulting from the inspection of an area containing no cracks.

Why would a noise distribution be bimodal? Such a distribution could result from a mixture of two distributions. (In figure 1, the noise distribution represents a mixture in the ratio of 9:1 for the two components.) There are several possible reasons that mixtures might arise. For example, an inspector using an inspection technique that is designed to inspect around fasteners inadvertently picks up a signal from the fastener edge. He is unaware of doing so and the process is a random one that is occurring about one-tenth of the time. The result is a mixture for the noise distribution.

A second example that would lead to a mixture is one in which there are physical differences between inspection items. That is, one-tenth of the inspected items has a condition that generates the elevated signal. An example might be a different fastener material or sub-layer material that is not easily recognized by the inspector.

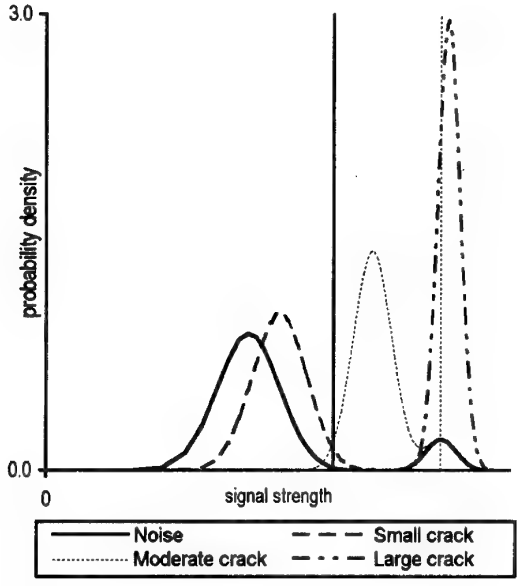


Figure 1. Signal Distributions for Hypothetical Inspection.

The signal distributions for the three crack conditions assumes the same mixing, but in the one-tenth of the time that the mixture condition is realized, the observed signal will be the maximum of the crack alone signal and the mixture signal. Using signal theory detection concepts [8] and the threshold as marked in figure 1 by heavy vertical line, one sees that the above model will result in a false call rate of approximately 0.10. The rate of detection for small cracks will be slightly higher than 0.10 and a model of the form of equation (3) is a good candidate for modeling the POD. (Note, we have not addressed the p_m parameter of that model, but arguments similar to the above would apply. The only difference would be that the resultant signal distribution for a crack of any length would be a mixture of a normal signal with a positive probability of being zero for conditions that result in a loss of signal.)

In fitting equation (3) to the above example we get an appropriate F( ) distribution controlled by the parameters μ and σ, $p_h = 0.10$ and $p_m = 0$. From the density curves of figure 1 it is clear that the function F would be close to 1 for both the moderate and the large cracks. However, it is decided that the decision threshold should be set higher in order to control the false call rate. The decision threshold is then set at the lighter vertical line. At this threshold the detection rate for the moderate size crack is no better than the false call rate. The detection rate for the large crack is only a little better than 0.50. It should be clear that the F( ) distribution derived using the first inspection decision threshold is vastly different from that that would be needed to model the second decision threshold.

The above example demonstrated that if the only control on the false call rate was to change the decision threshold, then it would be wrong to consider the F distribution as a POD model in the sense of equation (1). However, one example of a possible reason for the underlying mixture was procedural mistakes being made at random by the inspector. If this were the case and the cause could be removed through retraining or procedural changes, then the original F( ) function would reflect an achievable POD.

The above discussion points out that the mathematical form for POD given by equation (3) can fit multiple situations. To dissect the model further requires an understanding of the factors that lead to the variations of inspection. Thus, there is a role for DOE to help characterize inspection processes in the field.

4.2 Model Sensitivity and Comparisons

We illustrate the use of equation (3) to model POD with data taken using eddy current equipment to inspect for cracks in an inner layer. The work presented here comes from ongoing work sponsored by the United States Federal Aviation Administration at its Airworthiness Assurance NDI Validation Center in Albuquerque, New Mexico.

There were 98 flaws ranging in length from 0.25 mm to 14.8 mm. Most (54) of the flaws were in the range of 1 to 3 mm. In addition there were 260 non-flawed fastener sites that were inspected. Figure 2 shows the two-parameter lognormal fit to one set of detection data. An interesting characteristic of this fit was that the 90% detection crack length is estimated to be 11.8 mm. However, all 12 cracks that exceeded 3.5 mm in length were detected. There is a clear indication of a poor fit.

The fit using maximum likelihood estimation for the parameters of equation (3) of the 98 flaws is also shown in figure 2. The F( ) function of that equation is the lognormal cumulative distribution function. The fit now rises rapidly from the p_h estimate of approximately 0.27 to 1.0 between 3 and 4 mm. The fit effectively described 3 regions of the data. There were 19 detects in the 73 cracks below 3 mm in length, 7 detects in the 15 cracks between 3 and 3.9 mm and 10 detects in 10 cracks above 3.9 mm in length.

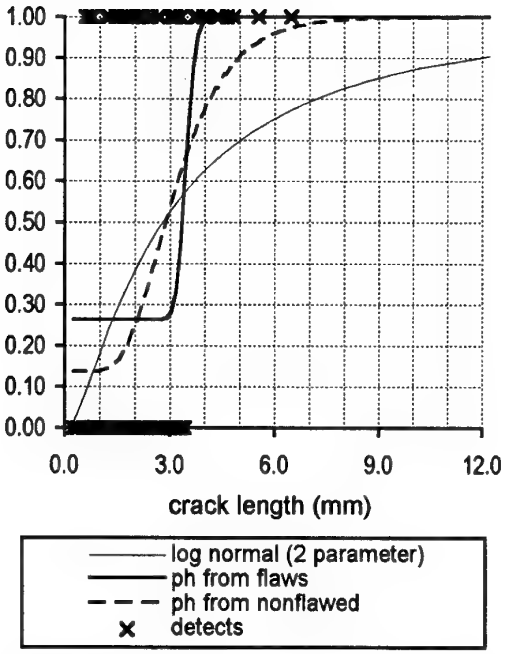


Figure 2. PODs from Different Model Assumptions

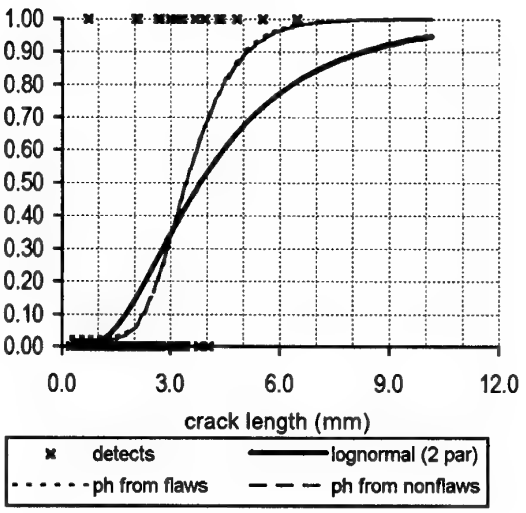


Figure 3. PODs from Different Model Assumptions-Inspector 2

If the parameter p_h reflects a false call rate then one could argue that it is best estimated by the rate of false calls made in the non-flawed population. For the inspection of figure 2 there were 35 false calls in 258 opportunities for a rate of 0.136. Using this estimate of p_h and then estimating the parameters of F( ) by maximum likelihood results in the third curve given in figure 2. We see that although the p_h is about half of that estimated within the flaw set, it still allows a substantial change in the upper end of the curve.

Figure 3 shows the same three fits given in figure 2, but to another inspectors data taken on the same test specimens and using the same equipment. The initial two-parameter fit results in an estimate of the 90% detection rate of approximately 8.3-mm. The p_h parameter is estimated from the flaw data as 0.029. The estimate from the non-flawed inspections is 0.019. The POD curves from these cases are very similar for cracks greater than 2 mm in length (detection rates above 0.06).

In the inspection of figure 3 there were a total of 51 cracks smaller than the second smallest crack detected. Notice that the smallest crack detected is separated from the rest of the detected cracks. By including p_h in the model we remove the influence of this small crack detection on the higher detection rate portion of the curve.

It is well known that with moderate amounts of data an extreme binary data point will influence the estimation of the scale parameter more than it will the location parameter. The above examples show that the parameterization of equation (3) removes some of the influence of small detections on the estimates made for detection rates of larger cracks. This parameterization is thus an effective way to gage and treat rogue points [9].

There is a drawback to using equation (3) and fitting the parameters by maximum likelihood methods. That drawback is that there are many local maximums for the likelihood equation. For example, consider any two adjacent crack lengths, a_i and a_{i+1}, when the data are ordered smallest to largest. If the location and scale parameters for F( ) are chosen so that almost all of the function change occurs between points a_i and a_{i+1} then the model reduces to estimating the parameter p_h by the proportion of detects in the population of cracks less than a_i and estimating p_m by the proportion of misses in the population of cracks greater than a_{i+1}. This solution will be a local maximum for the likelihood function. Thus care has to be taken in maximization search routines to conclude that a global maximum has been found.

4.3 Crack Length Distributions

In the data examples shown in figures 2 and 3 the two parameter lognormal fits resulted in estimates for the 90% detection crack length that seemed too large in view of the fact that no cracks of that magnitude had been missed in the inspections. However, there were

not very many larger cracks in the test specimens. Here we briefly examine whether the noted behaviors are likely due to the relatively few large cracks available.

Figure 4 shows the detection data from a third inspector of the same data set and conditions as presented in figures 2 and 3. The heavy curve in figure 4 is the POD estimated using the two parameter lognormal distribution. Using the model of equation 3, p_h is estimated to be 0.018. However, there is very little change in the likelihood and the effective change on the POD curve is minimal.

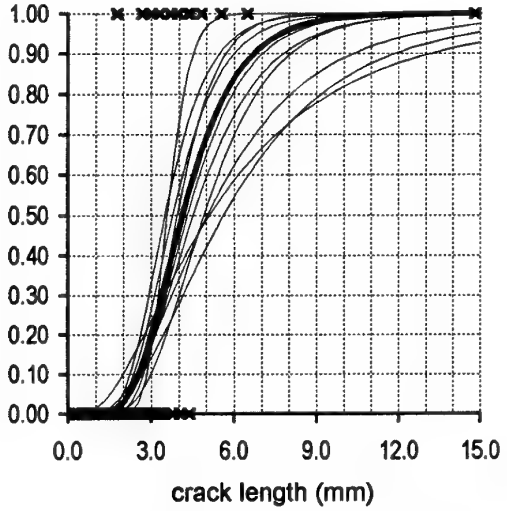


Figure 4. Assumed POD and 10 Random fits

We assumed that the lognormal POD fit to this data is, in fact, the true POD. We then generated 10 simulated inspections on the set of 98 cracks using this POD. For each of the simulated inspection outcomes we estimated the POD using the lognormal function. These ten estimated POD curves are also given in figure 4.

The ten curves cluster around the "true" POD curve. But the curves exhibit different variation in different regimes. The range in the estimated crack lengths for 90% detection is approximately 9 mm. The range in the estimated crack lengths for 10% detection is just a little over 1 mm.

Only one of the ten simulated inspections resulted in a positive estimate for p_h when fitting equation (3). It was a small value that did not affect the rest of the curve. However, five of the simulated inspections resulted in fits from equation (3) with positive estimates for p_m. In all five cases the estimate for p_m exceeded 0.10 and therefore 90% detection would never be achieved in these estimated PODs.

There were 8 cracks above 4 mm in length. The probability of detection associated with 4 mm is about 0.5. The relative scarcity of cracks in the 0.50 and above detection range for the 10 simulated inspection results is responsible for the variation in the upper portion of the curves (or in positive estimates of p_m when the model of equation (3) is used). Curve fitting to binary data is a form of regression. The curves of figure 4 show the variation that results when the fitted curves go beyond where data exist.

To illustrate the effect of the crack distribution on the estimated curves we assume a crack distribution chosen to be uniform in the log scale for the interval from the 0.01 to the 0.99 detection point of the same POD assumed in figure 4. Ten simulated inspections on this new data set were generated. Curves fit to each of them are shown in figure 5.

In figure 5 the ten curves still cluster around the "true" POD, but now we see a more uniform distribution for all detection levels. The range in the estimated crack lengths for 90% detection is about 2 mm, as it is for the 10% detection level. In all ten simulated curves the parameters p_h and p_m of equation (3) were estimated to be zero. This is as it should be since we assumed a true POD that was characterized by only the two parameters of the lognormal distribution.

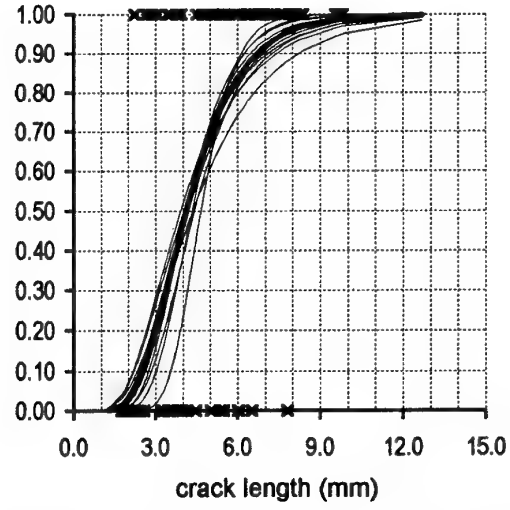


Figure 5. Assumed POD and 10 Random Fits - Cracks Distributed Uniformly in Log Scale

The results of the simulations shown in figures 4 and 5 demonstrate that the use of equation (3) with its four parameters needs to be evaluated with respect to the crack distribution. The lack of sufficient cracks at the lower or upper detection rates of the true POD will result in the variation of those rates being manifested as positive estimates of p_h and p_m. In other words, when using equation (3), positive estimated of p_h and p_m need to be interpreted with consideration of the crack distributions for the data.

Returning to the fits of figures 2 and 3, recall that both of the inspections resulted in estimates of p_h that were positive. The analysis with respect to the crack distribution implies that there were sufficient cracks at the lower end of the curve so that we believe the estimates are reflecting an aspect of the inspection process rather than reflecting uncertainty driven by the crack distribution.

5. ACKNOWLEDGEMENTS

The work reported here was partially sponsored through contract with Science Applications International Corporation/Ultra Image in support of a Warner Robins Air Logistics Center Program. The US Federal Aviation Administration Technical Center also provided support through the Airworthiness Assurance NDI Validation Center in Albuquerque, New Mexico.

6. REFERENCES

1. Spencer, F.W. and Schurman, D.L., "Reliability Assessment at Airline Inspection Facilities, Volume III: Results of an Eddy Current Inspection Reliability Experiment," DOT/FAACT-92/12, III, May 1995.
2. Mullis, R.T., "C-141 Spanwise Splice Advanced NDI Method," Airframe Inspection Reliability under Field/Depot Conditions, NATO RTO Workshop, May 1998, Paper 19.
3. Berens, Alan P., "NDE Reliability Data Analysis," Metals Handbook, v. 17, 9th ed, ASM International, 1988.
4. Annis, Berens, Bray, Erland, Hardy, Herron, and Hoppe, Proposed MIL-STD (1823) Non-Destructive Evaluation System Reliability Assessment, AF Contract F33615-81-C-5002, August 1989.
5. Berens, A.P., P.W. Hovey, R.M. Donahue, and W.N. Craport, "User's Manual for Probability of Detection Software System (POD/SS)," UDR-TR-88-12, UDRI, Dayton, Ohio, January, 1988.
6. Spencer, F.W., "Detection Reliability for Small Cracks Beneath Rivet Heads Using Eddy-Current Nondestructive Inspection Techniques," DOT/FAA/AR-97/73, to be published.
7. Fahr, A., et al, "POD Assessment of NDI Procedures Using a Round Robin Test," AGARD-R-809, January 1995.
8. Swets, J.A. "Assessment of NDT Systems - Part I: The Relationship of True and False Detections," *Materials Evaluation*, 41:1294-1298, 1983.
9. Hyatt, Kechter, and Menton, "Probability of Detection Estimation for Data Sets with Rogue Points," Materials Evaluation, November 1991, pp. 1402-1408.

A Systematic Approach to the Selection of Economic Inspection Methods and Intervals

S.H. Spence
British Aerospace (Operations)
Military Aircraft & Aerostructures
Warton Aerodome, W310C
Preston, Lancashire PR4 1AX, UK

ABSTRACT

Fatigue related inspections are required when the safe life of a structure is less than the target service life (as a result of shortcomings in design, changes in usage, etc.) or where structural integrity support by inspection has been identified by a damage tolerance analysis. The increase in life extension programmes arising due to constricting defence budgets is leading to an increasing dependence on inspections. Under these circumstances the structural integrity of a fleet or individual aircraft is safeguarded by inspection for fatigue cracks.

The majority of fatigue cracks in airframe structures occur at fastener holes. This work, therefore, specifically considers the inspection and repair of fastener holes. The inspectable crack size, inspection interval and cost are interdependent. The smaller the crack inspected for, the longer will be the period of growth to reach a maximum acceptable size. However, the associated preparation, inspection and down time costs will be greater. Further, for a given inspection technique, the probability of detection will be lower for a smaller crack and the chances of a false call will be higher. This paper discusses the factors which must be considered when selecting inspection techniques and determining the associated inspection periods. By optimising the inspection process, life-cycle cost benefits can be realised without compromising structural integrity. A schematic approach is detailed in which a balance may be struck between inspection effectiveness, required inspection interval and the associated costs. This will enable the end users to determine the most economic maintenance programme provided that the effectiveness of potential inspection techniques can be sufficiently quantified.

INTRODUCTION

Military aircraft are designed to meet set requirements in terms of life and reliability. In order to maintain this reliability with use, certain inspections are required to detect damage. This damage may be of various forms such as battle damage, stress corrosion, accidental impact and fatigue. There are two main types of structural inspections:

- General scheduled inspections,
- Specific fatigue-related inspections.

General scheduled inspections look for anything non-standard, for example stress corrosion, damage to protective coatings, wear and loose fasteners.

Fatigue related inspections are defined following specific arisings on major test structures or in service where the safe life or safe crack growth life is less than the target service life (as a result of shortcomings in design or changes in use and/or environment etc.) or in the case of life extension programmes. Under these circumstances the structural integrity of aircraft is safeguarded by inspecting for cracks at specific locations and at intervals frequent enough to provide an acceptable probability of finding a crack before it reaches a size where the residual strength is reduced below acceptable levels or where repair is no longer economically viable. Such practice is also required where structural integrity support by inspection has been identified by an initial damage analysis. This paper is concerned only with the fatigue related inspections.

Significant costs are associated with all inspection techniques but the actual level of cost is dependent on the particular technique used, (i.e. a single unaided visual inspection will be less costly than the use of rotary eddy currents), degree of preparation and refit necessary, down time and inspection interval. Whilst it is imperative that the reliability of the aircraft is maintained, that is the risk of failure is held at an acceptably low level, the costs of ownership must also be minimised. To this end, the inspection programme must be designed for reliability and minimised costs. There are many inspection techniques available including visual, liquid dye penetrant, ultrasonic scans, eddy current and magnetic particle. Different techniques have different merits, such as surface or subsurface detection, sensitivity, low cost and accessibility. Where more than one technique is suitable and available there may be a trade-off to be made between inspection effectiveness and interval. By optimising the inspection process, life cycle cost savings can be realised without compromising structural integrity.

Before a cost effective inspection philosophy can be defined, it is necessary first to gain an insight into the trade-offs available. The effectiveness of a technique is a measure of its ability to detect small cracks whereas the efficiency of a technique is a cost dependent quality. If the crack size inspected for, a_D, is reduced then the period for the crack to grow from a_D to a maximum acceptable size is increased. If this is brought about by the use of a more effective technique then the costs per inspection may increase due to increased preparation, inspection and down time. On the other hand, the reduction in a_D may be achieved by reducing the Probability of Detection (PoD) for a single inspection made using the given technique. This will, in turn, lead to an increase in the number of inspections required and in increased chance of a false call and the associated costs. Further, the overall costs will increase with the number of inspections required.

Three key requirements to enable a trade off study to be made for a given inspection programme are: Effectiveness of the relevant inspection techniques; risk reduction level required and; cost information.

DETECTION RELIABILITY

It should be noted that this and the following section reflects the views of the author and British Aerospace, Military Aircraft. Some issues, particularly with reference to PoD and the philosophy for determining the inspection interval are not

Paper presented at the RTO AVT Workshop on "Airframe Inspection Reliability under Field/Depot Conditions", held in Brussels, Belgium, 13-14 May 1998, and published in RTO MP-10.

in conformity with the Defence Standard 00-970, Leaflet 201/3.

If aircraft reliability is to be maintained above a minimum acceptable limit via inspections, it is necessary to quantify the reliability of such inspections in order to evaluate their contribution to structural integrity. The sensitivity and reliability of inspection is best characterised by a probability of detection curve (PoD plotted against crack length) (1).

If a feature of a component is inspected for the existence of a crack there are four possible outcomes of that inspection. If there is no crack at the feature inspected, either a correct non-indication will be recorded or an incorrect indication of a crack, known as a false call. False calls must be minimised since they can lead to unnecessary, costly repairs/ replacements. If, however, a crack of detectable size is present at the feature, either a correct indication will be recorded or an incorrect non indication will be registered, termed a miss.

If it is assumed that the outcome of an inspection process is independent of the operator, each technique will display a characteristic reliability which is dependent on crack length. This can be expressed as a probability of detection. The PoD is the number of correct indications as a proportion of the total number of opportunities for detection. Thus, for each technique, crack, material, geometry and environment combination, the relationship between crack length and PoD can be illustrated by a graph such as in Figure 1. It is common for the 95% confidence level to be used and the description of detection probabilities is no exception (1,3,5).

Inspection philosophies are often based on defining the detectable crack against the 90% PoD level (1-3,5). In order to obtain such data, a large sample is necessary to minimise differences between the mean and the 95% confidence curves. This is a very costly process and has typically been addressed by the "round robin" type approach (2,3-6). A number of these programmes have been undertaken as concern grew that the commonly quoted detection abilities for various techniques may have been somewhat optimistic (5). These programmes are a valuable source of data but are not without compromise (for example, differences between laboratory inspections of coupons and in-service inspection of components).

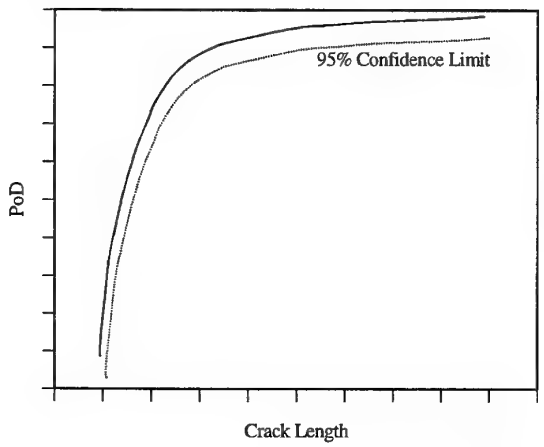


Figure 1. Variation of probability of detection with crack length including 95% confidence level.

INSPECTION PHILOSOPHIES

A number of inspection philosophies exist within the world civil and military aircraft industries (7-12). Inspection requirements which in the past were based on service experience and engineering judgement are now related to damage growth.

Inspection programmes are generally determined by defining the period of crack growth at a nominal risk level and then ensuring that sufficient inspections are conducted over this period to ensure an acceptably low level of risk associated with aircraft usage. The crack growth curve can be defined in a number ways (10-12). The most straightforward, in terms of calculation and data required, is the deterministic approach which employs the mean growth characteristics to calculate a growth curve from an assumed initial crack size. This curve can then be factored, in terms of life, as required. At the other end of the scale, a full probabilistic approach can be adopted involving the use of random variables to account for variations in characteristics such as initial damage or initiation periods, crack growth behaviour and fracture toughness values. Such analyses will produce predicted growth curves with an associated probability (12). Other approaches adopt a mixture of these two concepts. An example of this is where the initial fatigue quality is described by the Equivalent Initial Flaw Size, EIFS, concept and subsequent growth is modelled within a deterministic framework (10,11).

Within the EIFS concept the assumption is made that initial flaws, in the form of inherent material defects or handling/manufacturing danage, are present right from the onset of usage. Equivalent initial flaw sizes can be back-calculated from monitored crack growth and from total fatigue life data where a stress intensity factor solution exists and the crack growth rate behaviour for the material in question is sufficiently characterised. The EIFS is the calculated crack size which would result in the known final crack size after the known fatigue loading history. If a sufficient number of such calculations can be made then a probability density function may be obtained of EIFS, often termed the Equivalent Initial Flaw Size Distribution (EIFSD). This provides a measure of the initial fatigue quality of the component.

Under a deterministic approach, design or minimum specification fracture toughness values should be used to define critical crack length or appropriate factors should be applied to the mean value. The final crack size for inspection period definition is the largest crack size which is acceptable due to structural integrity or economic reasons. This is the smallest of the following crack size considerations:

- Onset of rapid, unstable crack growth,
- nett section failure including any allowance for loss of section due to corrosion,
- unacceptable leakage or loss of pressure,
- maximum size which can be practically or economically repaired,
- onset of any other failure mode, such as buckling, induced by crack growth.

Care must be taken to ensure that an appropriate level of conservatism is achieved in the calculated growth rate curves by carefully balancing allowances for initial fatigue quality and variability in growth rates and toughness. The objective is to achieve a growth curve with an acceptably low probability associated with the existence of a given crack length at the associated level of usage. As in any life analysis, sensitivity

studies should be carried out to ensure that small changes in applied loads do not produce disproportionate changes in life

It should be noted that the selection of the initial crack size, and hence initial fatigue quality, merely influences the total predicted life and the threshold inspection period (provided the initial crack is less than or equal to the detectable crack size). However, the allowance for variation in crack growth rates and fracture toughness will significantly influence not just total predicted life but also the inspection period and interval.

Whichever technique is employed, the calculated growth curve represents life to failure, if no inspections are carried out, with an associated probability. The inspection programme is then required to maintain the overall risk of failure at an acceptably low level. Generally, the lower the risk level associated with the calculated growth curve, the lower is the reliance on the inspection programme to maintain stuctural integrity.

PoD Evaluation.

Ideally, each NDI technique would have associated with it PoD-crack length data for specific component geometry/environment combinations. From these curves a crack size detectable at a specific level of probability could be defined, a_D. It is normal to define a detectable crack size with an associated PoD of between 50% and 95%. This requires there to be more than one independent inspections during the predicted inspection period. Generally, the inspection interval would be such that the accumulated PoD from all the inspections identified in the predicted inspection period is at least 99%. Practically, the probability for each inspection is assumed to remain at the level associated with a_D. For example, two opportunities for detection are required where the PoD for each inspection is 90%. If PoD data are not available, it may be necessary to make conservative estimates.

Defining Inspection Intervals

Having established the component calculated growth curve via one of the techniques described earlier and the potential NDT techniques and associated detectable crack sizes, it is necessary to determine the inspection threshold (if applicable), period and interval (13). Ideally, this would all be performed on the basis of aircraft usage in terms of whatever unit served as the best indication of fatigue life consumption, Fatigue Index (FI) or Flight Hours (FH), depending on whether fatigue damage accumulation is mission or flight-time dependent.

The threshold period is that period from the onset of service use to that point by which the first inspection must be executed. These and other inspection parameters are illustrated in Figure 2. The threshold period is dependent on factors such as the detectable crack size and the type of structure to be inspected.

The inspection period follows on from the threshold period and is limited by the crack attaining the maximum acceptable size. The inspection interval represents the maximum allowable usage between inspections (13). It is determined by dividing the inspection period through by a factor, F, (2 for a PoD of 90%). This factor, F, is defined by the number of repeat inspections required to achieve the required accumulated PoD, e.g. 99%. To allow for greatest flexibility for the scheduling of inspections, particularly with regard to integrating maintenance services and fatigue related inspections, it is simply required that inspections must be conducted **within** the inspection interval rather than, necessarily, when specific FI or flight hours are reached.

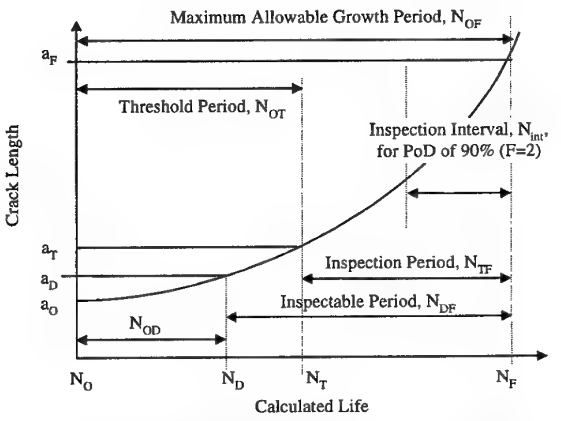


Figure 2. Calculated component growth curve including the inspectable period and inspection period, interval and threshold.

INSPECTION EFFECTIVENESS AND EFFICIENCY - THE TRADE-OFF.

Acceptance of the PoD against crack length data for quantifying the effectiveness on inspections is widespread but not universal. It is argued that the costs and resources required to establish meaningful PoD data are not practicable or justified.

It is also suggested that NDI selection and validation could be based on experience and limited testing. However, high, but ill-defined levels of confidence exist in the ability of each technique to reproducibly detect cracks of a size equal to, or greater than, a stated value.

This does not provide for the most suitable basis on which to make a trade-off between effectiveness and inspection interval because it does not quantify inspection effectiveness. As a result, less confidence can be held as to whether the cost optimised inspection programme still provides the same level of risk reduction as other slightly more costly options.

The more precisely the effectiveness and costs of a selected approach are known, the better are the cost optimisations and the more confidence can be enjoyed in the effectiveness of the maintenance operation resulting from the trade-off. Simply put, to make a valid trade-off, the effectiveness and efficiency must be well defined.

Minimisation of Cost

A convenient measure of the cost of an inspection programme is the maintenance-man-hour per flight hour (MMH/FH) or MMH/FI. The trade-off above aims to minimise MMH/FH costs without compromising reliability. With the knowledge of the general inspection strategy, the effectiveness and costs for each available inspection technique, it should be possible to identify a minimum cost option. A practicable, realistic view of the approach should be preserved in order to maintain structural integrity at a minimum cost.

Inspection Factor, F	PoD %	a_D (A) mm	a_D (B) mm	Inspection interval for A $(a_{cr}-a_T)/F$	Inspection interval for B $(a_{cr}-a_T)/F$	Cost /inspection (A & B)	MMH /FH for A %	MMH /FH for B %
1	99	2.5	1.25	1350	2350	1	0.074	0.043
2	90	1.25	0.55	1175	1500	1	0.085	0.067
3	80	0.55	0.3	1000	1133	1	0.1	0.088
4	70	0.3	0.2	850	962	1	0.118	0.104
6	60	0.2	0.15	641	700	1	0.156	0.143
7	50	0.15		600		1	0.167	

Table 1. PoD, inspection programme and cost information for Example 1, inspection techniques A and B, $a_i=a_D$.

Inspection Factor, F	PoD %	a_D mm	Inspection interval $(a_{cr}-a_T)/F$ [FH]	Cost/ inspection	MMH /FH Inspn. only %	Total Cost/ Inspection	MMH /FH Total costs %
1	99	2.5	450	1	0.22	50	11.1
2	90	1.25	392	1	0.26	50	12.8
3	80	0.55	333	1	0.3	50	15
4	70	0.3	283	1	0.35	50	17.6
6	60	0.2	214	1	0.47	100	46.8
7	50	0.15	200	1	0.5	100	50

Table 2. PoD, inspection programme and cost information for Example 2, inspection technique B .

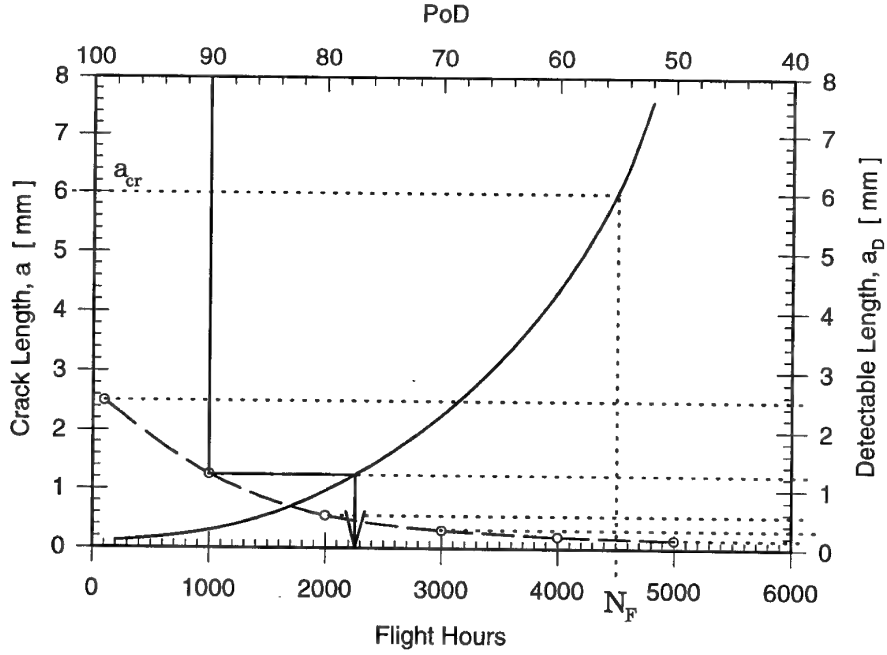


Figure 3. Calculated component growth curve (solid) and PoD-crack length curve (dotted). The intercepts of the crack length lines from the PoD curve with the growth curve, define the beginning of the inspectable period in terms of flight hours or FI.

The task of costing an inspection programme for a given strategy in terms of MMH/FI (or MMH/FH) is complex. The cost associated with the actual inspection of a given feature or features for a particular component can be estimated reasonably readily. Further, this will not necessarily vary to any great extent between inspection techniques. However, other costs involved such as those listed below can vary tremendously:

- Gaining access to the component,
- preparation of the area to be inspected,
- recovering the component condition (e.g. replacing protective coatings),
- replacing components or other equipment which were removed to gain access,
- down-time (non availability of aircraft),

The level of these costs depends not only on the inspection technique but also whether or not the inspection coincides with a scheduled maintenance service and if so which type of service (e.g. primary, minor, or major). Obviously, all inspections need to be scheduled such that they coincide with a maintenance service where at all possible otherwise severe cost penalties will be incurred. These costs clarify the requirement for a degree of flexibility in the inspection programme, hence the stipulation that an inspection must be made *within an inspection interval*, rather than at a specific flight hour or FI, to facilitate integration with a maintenance service. Cost benefits for components which require removal of other parts in order to gain access for an inspection will be particularly sensitive to integration of more major maintenance services during which the removal of the relevant parts may be required.

Using the inspection programme information (a_D, inspectable periods, inspection periods, intervals and thresholds) and an estimate for the cost of one inspection, the maintenance-man-hour/flight hour costs can be estimated for each technique deemed suitable for the particular component. For illustration, consider the calculated component growth curve (deterministic in this example), presented in Figure 3, and the inspection, PoD and cost information from Table 1. The PoD data are also included in Figure 3. The points where the crack length lines from the PoD curve intercept with the growth curve, define the beginning of the inspectable period in terms of flight hours or FI for each PoD level. For example, the figure indicates that for an a_D of approximately 1.2 mm and PoD of 90% the inspectable period commences at about 2200 FH. These figures are for illustration only but are representative: The component growth curve is representative of a fastener hole in a lower wing skin panel whilst the cumulative PoD data are representative of the rotary eddy current technique, here referred to as technique A. Data for a hypothetically more effective technique, B, has been created by factoring the data for technique A to give the information included in Table 1.

Example 1

To explore the cost implications, consider a typical plot of predicted inspectable period (number of flight hours to grow a crack from a_D to a_{cr}) against a_D, derived from Table 1 and illustrated in Figure 4. It can be seen from the figure that, for a given technique, A, if a smaller crack is chosen for detection at a lower associated PoD, then the inspectable period increases. However, to maintain the same overall probability of detection, the number of inspections to be made during the period will have to increase. It can be seen from the curve for technique A in Figure 5 that in this instance, costs increase with reducing a_D due to the increased number of inspections required. In this example there is a marked increase in costs associated with a reduction in PoD levels below about 65% (a_D~0.4).

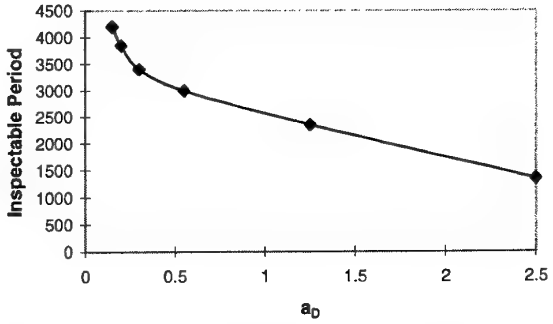


Figure 4. Dependence of inspectable period on detectable crack length, a_D, for technique A.

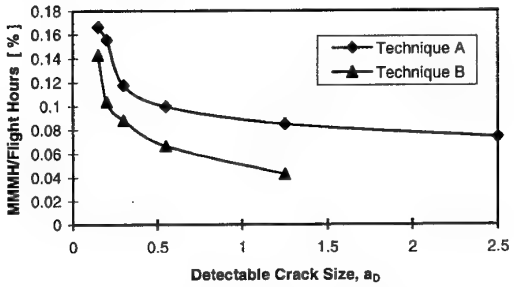


Figure 5. Influence of the detectable crack size, a_D, on cost in terms of maintenance-man-hours/flight hours for techniques A and B.

If a more effective technique were to be used, i.e. one with smaller values of a_D associated with equivalent PoD levels, then there may be cost savings as a result of the increased inspection period. This can be investigated by comparing the results for technique B which in this example has the same cost per inspection as technique A. The costs associated with inspection programmes employing each technique are illustrated in Figure 5 for comparison. Once again, for technique B, costs increase with the number of inspections but the costs for a given a_D are lower than for technique B. For example, from Table 1 it can be seen that the cost for a_D=0.55 mm has reduced from 0.09 to 0.07 MMH/Flight Hours (%).

If the cost per inspection were to be 25% more for technique B, then the cost would increase to 0.08 MMH/Flight Hours but this still compares favourably with the original cost of employing technique A. Incorporating technique B into the inspection programme would therefore realise a saving even at the higher cost per inspection.

The example so far has been simplified by only considering the cost of the inspection itself. However, much greater costs can be incurred due to preparation, access, refit and down time. If inspection intervals are very large such that inspections can be integrated with the more thorough maintenance services

whatever the technique used, then the additional costs of gaining access and subsequently replacing parts are less likely to be significant. Similarly, down time costs directly associated with the inspection should be minimal or non existent.

An additional area which could perhaps influence the overall inspection cost is that of preparation and protective coating replacement. For example, the dye penetrant technique requires paint removal whereas the eddy current method does not. Further, it may be possible to identify a trade-off where the necessity for fastener removal is dependent on the inspection technique employed (where the maximum acceptable crack size allows for both techniques). Leaving the fasteners in (if possible) would probably result in a relatively small inspectable period required for the less effective technique but with lower associated costs due to non removal of fasteners. However, whilst removing the fasteners will increase the cost of each inspection, it will result in a larger inspectable period.

On the other end of the scale, where inspection intervals are small, of the order of, or less than the primary maintenance service interval, then all the costs associated with each inspection technique will have a significant effect on the overall cost of inspection. In this case, it is possible that the minimum cost approach may be more marked. This would be particularly so if, for example, one technique lead to an inspection interval less than the primary service interval but an alternative technique provided for the integration of the inspection with the maintenance service. However, at this level costs may still be high if component removal is required to gain access to inspect the relevant component since the primary service is usually a relatively limited maintenance operation.

Not all inspections will require removal of parts to gain access in which case such considerations are less relevant. However, it is clear that the overall cost of an inspection programme which does require component removal for access will be particularly sensitive to integration with an appropriate maintenance service. The appropriate down-sizing of inspection interval and remapping onto the calculated component growth curve would be necessary in this instance and the costs estimated accordingly. This could lead to a somewhat iterative process for each inspection technique being considered. However, an estimate of the minimum attainable MMH/FH (or FI) should be achieved for each available technique. The inspection programme which gives rise to the lowest cost can then be employed.

Example 2

Consideration of costs associated with issues such as gaining access and down-time can be introduced into the model used in Example 1. In this second example case, the calculated component growth curve has a lower associated life. Table 2 displays the information for this new growth curve within the previous model with the increased costs displayed in Figure 6. Costs associated with preparation and refit for inspection and down-time are included in Table 2. The effect of these costs on the overall cost of ownership in terms of MMH/FH for the inspections, down-time, etc., is illustrated in Figure 7. The figure clearly demonstrates the marked increase in cost incurred as the PoD for crack inspection is reduced, and hence a_D with it, resulting in the inspection interval reducing to such an extent that inspections can no longer be integrated with suitable routine maintenance service.

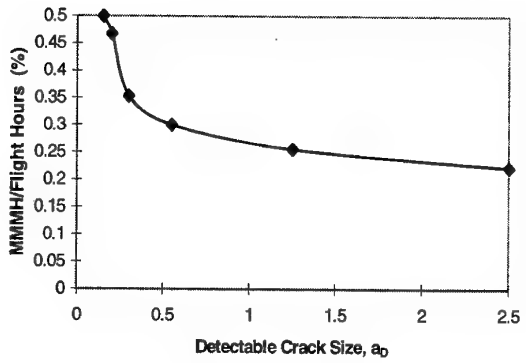


Figure 6. Influence of the detectable crack size, a_D, on cost in terms of maintenance-man-hours/flight hours for example 2.

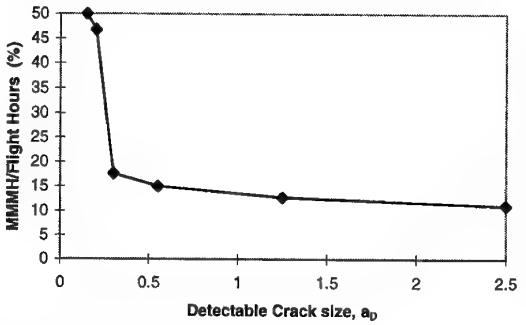


Figure 7. Influence of the detectable crack size, a_D, on cost in terms of MMH/FH example 2 where preparation and refit costs are included in addition to those for the actual inspections.

This model is still a simplification of the problem and it is at this stage that the iterative process begins for each available inspection technique. The routine maintenance services should be mapped onto the calculated component growth curve and the inspections adjusted to integrate with these service intervals where possible. The appropriate costs associated with preparation and refit for each inspection and the down time for each service category should then be associated with the relevant inspection and the cost calculations repeated as before. This process will finally highlight one technique and inspection programme to provide the minimum cost option at the desired risk level.

Finally, whilst it is clear that fatigue related inspections should be integrated with service intervals wherever possible, other less immediately obvious factors must also be considered. For example, consideration should be given to the environment under which the inspections are to be conducted. For instance, it may not be prudent to stipulate an inspection procedure which involves removing protective surface coatings in an open marine environment.

CONCLUSIONS

Methodology for identifying minimum cost inspection programmes has been discussed with a view to minimising the overall cost of aircraft ownership. Emphasis has been placed

on full identification of the risk of failure and probabilities of detection for reliability maintenance. However, it is realised that such detailed data may not be available. History shows that the inspection programmes have been based more on engineering judgement and the experience of expert engineers than on in depth calculations. Excellent service records provide testament to reliable design and inspection programmes. However, there is a move towards more defined quantification for fatigue related inspection programmes to provide cost optimisation without compromising reliability and safety. Further, the trend towards damage tolerant design philosophy in the military and civil industries and requirement for inspection through design in civil aircraft, emphasises the requirement for cost optimisation.

The trade-off process outlined in this work can be applied irrespective of the technique used to predict the component growth curve whether deterministic, probabilistic or o combination of these approaches.

In some cases, for existing aircraft, there may not be sufficient data to enable a trade-off study to be undertaken without a costly and perhaps lengthy study where the more traditional experience based approaches will be more appropriate.

The following conclusions can be drawn from the discussion on cost optimisation of inspection programmes and the examples employed:

1. For a given inspection programme a trade-off study is not possible to optimise costs unless an evaluation can be made of the effectiveness of potential inspection techniques.
2. Limited, if any, cost gains can be made from selecting a lower PoD and smaller associated detectable crack length, a_D, for a given technique.
3. Potentially, cost savings can be realised via the use of more effective inspection techniques.
4. The integration of inspections with routine maintenance service schedules plays a critical role in the optimisation process whereas inspection technique costs play a less significant role.

REFERENCES

1) Sampath, S.G., "Airframe Inspection Reliability". AGARD SMP Lecture Series on *Aging Combat Aircraft Fleets - Long Term Applications*, LS-206, 1996, pp 12-1 - 12-9.

2) Fahr, A., Forsyth, D., Bullock, M., Wallace, W., Ankara, A., Kompotiatis, L. and Goncalo, H.F.N., "PoD Assessment of NDI Procedures Using a Round Robin Test". AGARD report no. 809, 1995.

3) Simpson, D.L., "Development of Non-Destructive Inspection Probability of Detection Curves Using Field Data". Laboratory technical report: LRT-ST-1285, National Aeronautical Establishment, National Research Council Canada, 1981.

4) Bruce, D.A., Curtis, A.R. and Jackson, G., "A Statistical Investigation to Determine the Practical Reliability of N.D.T. Methods With Regard to the Implementation of Damage Tolerant Design Philosophy". Report no. BAe-WMD-RP-RES-NDT-000020, British Aerospace Defence, 1988.

5) Ludas, K.J., "Influence of Aging Aircraft Programs on the MD-11 Damage Tolerance Certification Process". Proc. 17th ICAF symposium, Stockholm, Sweden, 1993, pp 1167-1190.

6) Lockheed-Georgia Company, SA-ALC/MME 76-6-38-1, "Reliability of Non-Destructive Inspections", 1978.

7) Goranson, U.G. and Rogers, J.T., "Elements of Damage Tolerance Verification". Proc. 12th ICAF symposium, Toulouse, France, 1983.

8) Goranson, U.G., "Damage Tolerance - Facts and Fiction". Proc. 17th ICAF symposium, Stockholm, Sweden, 1993, pp 3-105.

9) Miller, M., Luthra, V.K. and Goranson, U.G., "Fatigue Crack Growth Characterisation of Jet Transport Structures". Proc. 14th ICAF symposium, Ottawa, Canada, 1987.

10) Yang, J.N., Manning, S.D. and Newman, Jr., J.C., "Assessment of Initial Flaw Size Technologies for Metallic Airframes". Report No. UAI-95-1, United Analysis, Inc. Air Force Subcontract No. 312138, 1995.

11) Manning, S.D. and Jang, J.N., "Probabilistic Durability Analysis Methodology for Metallic Airframes". Proc. 17th ICAF symposium, Stockholm, Sweden, 1993, pp 321-345.

12) Tisseyre, M., Plantec, J.Y., Beaufils, J.Y and Boetsch, R., "Aerospatiale Probabilistic Methods Applied to Aircraft Maintenance". Proc. 17th ICAF symposium, Stockholm, Sweden, 1993, pp 589-618.

13) Spence, SH., "Fatigue Related Inspections: The Trade-off Between Effectiveness and Inspection Interval". Report no. BAe-WSS-RP-RES-SOR-000317, British Aerospace, 1998.

THE EFFECT OF AIRCRAFT MAINTENANCE ON HUMAN FACTORS

MWB Lock
School of Industrial and Manufacturing Sciences
Cranfield University
Beds MK43 0AL UK

Summary
Many factors affect the performance of the human operator during the inspection and maintenance of aircraft. This paper highlights the basic problem of attempting to quantify these human factors. To ensure reliable task performance it is suggested that one should reduce these effects by ensuring operator comfort, both physiologically and psychologically rather than attempt to estimate probabilities.

Introduction
Many factors affect the reliability of human endeavours. The more exacting the task to be performed; the more critical is the need for knowledge of the factors affecting its successful completion.

Inspection of aircraft and the associated maintenance activities have to be among the most critical tasks that man undertakes and much observation and research[1-6] has been done in an effort to illuminate the problems caused by man's fallibility. The price of failure is often death.

Thirteen years ago the author and his erstwhile colleague, Dr John Strutt, were commissioned by the CAA to carry out a similar survey of structural inspection within the transport aircraft fraternity[1]. We were given a reasonably open-ended remit to report from the standpoint of uncommitted scientists. In the early nineties, following the 'Aloha' incident the author was asked to do a similar survey and to include Non-Destructive Testing[3,4,5].

This paper, therefore, comes from someone who has a limited experience working actively in this industry but who has spent many months talking to those concerned and watching them during their work times.

Our conclusions in 1985 ended by stating that progress would be made, not in big sweeping changes but in many small, seemingly insignificant, ways. This proved true. Great automated machines in aero-hospitals are still not to be seen assessing structures although NDT is making many inroads: inspection still relies, predominantly, on the human eye and the condition of the human behind it is of major significance

The burden on the inspection process from inception, through execution to final quality acceptance is very great. It is not difficult to imagine the task as being too complex for a human system to control and yet past experience shows it to have been remarkably reliable. How much this success is due to design and how much to providence is debatable but it remains a fact that very few past incidents (a better word than accidents) can be placed solely at the feet of the inspection or maintenance processes

Reliability is the final end-product of a whole series of activities of which an exceptionally important one is the Human Factor.

What Factors are Human?
Factors which are human-based can be arbitrarily divided two camps, Physiological and Psychological.

Physiological factors involve the usual five senses:
Sight, Hearing, Taste, Smell and Touch.
There are then secondary effects on those senses:
Glare, Noise, Cleanliness, Position, Comfort and Safety.

Psychological factors can be subdivided arbitrarily into:
- **Personal** Confidence and Health.
- **Imposed** Training, Feedback, Responsibility and Supervision.
- **Organisational** Management and its organisation of Terms & Conditions, Shift Systems and Bonding (claw-back of training costs).

How do these factors interact?
We are only too aware that 100% reliability is a pipe-dream. Consider the interaction of just a few factors on a simple inspection task. Consider, for an instance, the effects of hangar temperature or an external event such as a family row or bending the motor in the car park.

At comfortable temperatures and without the external event, the operator will attain a certain 'normal' level of reliability. Here, reliability is loosely defined as the probability, P_o, that a task will be completed satisfactorily. This will depend on all the other factors, human and otherwise and will most probably be less than 100%. If the temperature is lowered, there will come a point when the operator's reliability drops.

The overall effect will be complex but could be written:

$$P = P_o (1 - f(T)) \quad \ldots\ldots \quad (1)$$

where P is the probability at a particular temperature, all else being equal and f(T) is a function of temperature. f(T) is not single valued: its form is unknown and certainly not simply proportionate.

Further changes to f(T) arise if the operator attempts to alleviate the discomfort by donning a coat and is then, perhaps, less able to fit into an access hole during a task on the wing or if he puts on gloves which may then render the handling of a NDT probe less effective. Any one of a dozen such sub-factors will alter f(T), perhaps subtly, perhaps not. Certainly, if the operator tries working on a night shift at 5°C (42°F) as did the author on one occasion, the act of donning coat and gloves is very conducive to reliability improvement. But when

Paper presented at the RTO AVT Workshop on "Airframe Inspection Reliability under Field/Depot Conditions", held in Brussels, Belgium, 13-14 May 1998, and published in RTO MP-10.

does the increase due to comfort offset the reduction due to obstruction?

Equation 1 now changes according to the interrelationships of the new factors depending on how they interact and to which set of statistical formulae you subscribe.

perhaps $P = P_o (1 - f(T).g(C))$ 3
or $P = P_o (1 - f(T).h(G)$ 4
or $P = Po (1 - f(T) + (g(C))$ 5
etc. etc..

Here each function f(T), g(C) and h(G), relating to temperature, coat and gloves, is non-linear and will change according to the operator and the task attempted: eg gloves affect area checks less than probe handling; coats, the reverse.

Increase in temperature has its own problems; availability of water, fans, sweaty grip and so on.

NOW, consider the effect of the external event; the family row or the bent car. If the operator broods on this it may lower his (or her) concentration and we might express it as:

$$P = P_o (1 - p(E)) \text{ ..... } (6)$$

How does this new factor affect the effect of temperature change already considered. Does the operator don his coat earlier and increase his reliability with respect to temperature or does he notice the addresses exchanged with the accidentee in its top pocket and so reduce it?

The final probability due to these events is not clear. Even for two reasonably straight-forward events the possibilities increase rapidly and the equations get out of hand eg:

$P = P_o (1 - \{or +\} f(T) . \{or + \{or-\}\} g(C) or H(G)$
$\{or both\} . \{or +/-) p(E))$

That is $P = P_o (1 - fn(T,C,G,E))$ where the interrelations of the separate factors is poorly understood, and most probably indeterminable.

Does it matter when the effects are so small?
If we considersome scaling functions, such as f(T) above, to have values of 0.02 (98% probability of completing a task successfully) then we might combine 10 of them as follows:

If they were completely independent factors one could contemplate probabilities of any thing from zero, where the factors cancel each other out (very unlikely), to the worst case where the factors positively interact in the same direction, ie we could simply add the probabilities eg 10 times 0.02 = 20%. A realistic figure is somewhere between.

Factors, one at a time
As mentioned above, the factors may be divided into two types, physiological and psychological. There follows a brief discussion of some of the problems which the author saw in his site visits and which arise with both types.

Can you see it?
To locate defects and assess corrosion is not solely a case of having sharp vision able to focus at both near and far distances but also of having a broad perceptual capability. The inspector needs to be able to perform close inspection inside restricted areas where the eye can often get no farther than six inches or so from the subject and then needs to be able to make a fast but comprehensive scan of several feet of surface.

Drury[7] summarises the dependence of crack detection with the angle of the eyeline, showing that visual acuity reduces by a factor of two as the angle of viewing increases from 20 to 40 degrees. He also emphasises the relevance of the visual lobe, the area around the line of sight within which a defect can still be detected and the consequent trade off between viewing time and detection probability.

It is a universal requirement for inspectors to have good vision, with glasses if necessary, for employment as a visual inspector and this is usually checked in the initial medical examination, taken when entering a company. Second rate vision is not good enough.

Subsequent examination is not a normal requirement in the UK, and so the difficulty lies in how to deal with the situation of worsening vision during employment. Those who rely on vision to perform their tasks properly need to have it well-defined in the first place and continually re-assessed.

A question arises as to whether it is appropriate for the employer to pay for the maintenance of the inspector's vision: one can see both sides of the argument.

i. The inspector should be responsible for his being able to perform the tasks
ii The employer should treat this as another of the tools required in the inspection process.

It has to be accepted that sight usually worsens very slowly so that the inspector is not aware of a sudden change. This makes it all the more important that tests are given. Professional bodies such as the Association of Optical Practitioners would be able to suggest a set of vision requirements.

There seems to be a reluctance to make staff redundant if they are unable to satisfy visual requirements. Redundancy would be the norm if they were to become incapable physically or mentally, through accident or otherwise.

Some form of compensation may be required. A similar circumstance is met, through an insurance scheme, for pilots who fail their fitness tests. An inspector's eyesight is a lesser problem and unlikely to be subject to the same long-term financial implications. There are, though, similar safety considerations and some form of insurance might be considered for those whose safe working life reaches a premature end.

Visual Inspection Aids. These act as human sense enhancers providing improvement by amplifying, magnifying, reflecting etc. These are usually quite rudimentary: a torch (US: flashlight), a hand-lens and a mirror on a stick.

Torch The requirements for a torch are basically simple.It needs a uniform beam of adequate brightness in a suitable case.

Uniform, so that any surface smudges and colouring which can act as good tell-tales of leakage, cracking etc do not get lost in the soft marbled effect of so many surfaces seen.

Adequate brightness, because we are not looking for a keyhole in the front door but for a 5mm long, hair line crack which is just peeping out of the muck around a flange where the cleaner has not quite reached and which has been even further smeared by the inspector's rag and reflected by a dirty mirror.

A suitable case, because we do not want one that softens in hydraulic fluids and corrosion inhibitors. It needs to be able to withstand a drop from halfway up the tailfin onto a concrete floor and possibly being run over by a forklift truck.

You cannot get a torch to these specifications for £1.55p from your local tobacconist. The situation is better now than ten years ago with the availability of a range of torches with tubular metal cases and variable focus beams which are easily adjusted to provide a small diameter intense beam or a wider, general area light. Even the smallest are bright enough for close-up work where space is at a premium and are admirable for taping onto a stick-mirror's handle.

Companies have not been eager to supply such quality torches due to alleged theft problems; how much more this applies to the inspector himself! Incidentally, in the USA, all flashlights used for inspection have to be underwriter approved.

Stick-mirror Commonly recognised as a dentist's mirror, this is the other half of the inspector's armoury. Yet many are still cracked, have loose swivels and are losing their backing. The maintenance of this item is often poor, relying frequently on elastoplast to a greater extent than its owner.

Hand lens The hand lens is another neglected tool being habitually scratched in the centre due to its convex nature. As this is the area which contributes most to the definition of the magnified image it is not acceptable in that condition.

Lighting In general, lighting in hangars is acceptable and daylight phosphors combined with ambient daylight provide a good general working level. The problems start at the inspection site when the inspector is actually looking at something. Several things may disrupt this good background lighting level.

Obviously the background level is largely obscured when the inspection takes place under the aircraft, inside an undercarriage bay or in a wing tank. Subsidiary lighting is often provided but rarely in sufficient quantity to satisfy the needs of all the work-force. Too often a job has to wait because lights are not available or are of the wrong type eg cannot be angled as required or have cables which are not long enough. Don't even think about trying to get another extension lead from the stores; there aren't any. Many inspectors carry their own out of exasperation.

Lights are installed in the floor in some cases and this can be useful although they are easily obscured. They are also installed in staging, either suspended from the level above or on lamp-posts. These auxiliary lights need to be quite powerful to be useful and this can lead to difficulties. In many installations these are single incandescent bulbs which have major disadvantages compared with striplights. The light from these decreases with the square of the distance (whereas striplight illumination varies directly with distance). This means that the light gets dimmer, faster, as you get farther away and point sources are therefore less flexible in use. The other is that they are distracting sources of glare.

Glare is a problem from several other sources. Bright sunlight from windows or rooflights is one. Standing atop one aircraft in its hangar the author was only a few feet from bare glass rooflights. The glare from the aircraft crown was so bright that it was extremely difficult to see the eddy current display 'scope. Similarly, glare from structural reflections around an inspection hatch can dazzle.

Another source can be the sudden flash from pinpoint sources due to roof-hung or subsidiary lighting. These bright flashes can render the inspector's vision far below par for a period of many seconds as anybody who has lain under a car trying to adjust the fan belt will be aware.

During this time an inspector is unable to perform with the normal reliability. The tendency when this happens is to screw up the eyes or rub them for a few seconds and to continue with the task. This is not sufficient to reinstate vision. What is required is an established method for assuring inspectors that their vision has returned to normal. Even waiting for a fixed time (longer than you'd think) would do.

The IES code contains ASA standards on lighting levels for various types of work as measured by light meter etc. An in-house comparison of recommended levels for inspection purposes with the actual light levels at inspection sites would be informative..

Does it smell?

As a function of the workplace, smells are not major or even frequent problems and, in themselves, do not affect reliability directly. Most complaints occur when the inspector is working around toilet and kitchen areas or in connection with fuel tanks and diesel engine fume.

The problem in all the areas leads to the same effect. The inspector wishes to get away from the situation as quickly as possible. It is not a question of the inspector

being too delicate and smell can often be the symptom of a failure in sealing or protection and as such is a useful fault indication.

In fuel tanks the condition couples with the uncomfortable regime encountered in restricted height amid tortuous structure. In the oil industry it is quite common to add deodorants to the more sulphurous liquids and gases and there is a well established industry to advise the operator if asked. These substances could be administered during the cleaning cycle to the benefit of all those about to work inside the tanks.

Making the environment smell less nauseous should not be viewed as a cosmetic exercise but as an aid to inspector comfort which leads to enhanced reliability.

Is it too noisy?

Noise is a **real** hazard in the hangar. It stems from generators, air compressors, air-operated and other power tools and APUs.

Good hearing, like smell and taste, is a handicap in most situations as far as its contribution to reliability in structural inspection is concerned; indeed an impaired auditory sense could be an advantage in many hangars visited. We should not routinely try to deafen people with rivet guns, APUs and compressor noise. We must avoid this direct interference with the inspector who is trying to tap-test for delamination, or is joggling a joint to detect wear in journals etc.

For the majority of inspectors, noise has a significant effect on their efficiency by reducing concentration and creating tension although it has been found to increase short-term vigilance, especially when combined with sleeplessness (the cussedness factor, Hockey[8].

Silencers are available for many power tools but are rarely seen. They are often of cumbersome design making the tool more difficult to handle but engineers should be encouraged to use them. They are best obtained at the time of purchase as it is, unfortunately, only too common for 'non-essential' purchases to be refused or continually delayed at a later date.

While the odd loud noise is temporarily distracting it is the continuous compressor noise or continual repetitive noise such as panel removal which causes the inspector to stop in the middle of a task.

Extended use of noisy tools might be kept to fixed periods during the shift. A **'noise schedule board'** is a viable proposition and could be continually up-dated by those who are able to plan their essential noise timetable and referred to by inspectors about to embark on a lengthy inspection.

This would reduce the possibility of missing a part of a repetitive inspection due to the inspector stopping because of continuous noise, the task being resumed later at a different point. Although only a small risk, it might prove significant while performing a large area check
or during the eddy current NDT on a lap joint or even tap testing a number of composite panels.

For the latter, the coin is still the most used tool for tap-testing and with the advent of the gibbously heptagonal 50p piece, tap-testing has entered a new era with the choice of two radii. It is also easily obtained, rugged, wholly reliable, keeps its value and is simply replaceable: the perfect tool.

However, it is worth noting here that while definition of the minimum area of delamination permissible is a feature of definition in the MM, the mode of tap-testing is frequently not. A simple experiment on an old panel will convince one that a delamination does not reveal itself to a tap more than an inch or two away.

For a panel measuring 3' x 2' the minimum number of taps to ensure soundness would be 216 at corners of a 2"mesh but the author has rarely seen this many; most of them are done around the edges with a few in the centre, for luck. Even allowing for previous experience of similar panels this is insufficient. This is an area where the inspector needs more information.

Noise levels must be reduced where possible by the use of mufflers for equipment, putting noisy equipment outside the hangar where possible and if all else fails by insistence on the wearing of superior quality ear-plugs or muffs. Adherence to the HSE guidelines is not enough. We are not trying to prevent deafness but to improve reliability by improving the inspector's concentration.

Can you feel it?

Much inspection is effected by touch rather than by looking. Bearings, universal joints etc are tested by working them and 'feeling' for play rather than 'looking' for it. Visual inspection is closely linked with the sense of touch. While looking at cables or surfaces, the inspector is also seeking evidence of incipient fraying or slight irregularity by running fingers along them. The tactile sense is therefore not only a primary inspection tool but with visual feedback is also complementary to visual inspection.

It is diminished substantially if the wearing of gloves is necessary due to cold or excessively dirty conditions which can occur if the inspection is being performed out on the ramp or down-the -line.

Related to touch is the sense of balance or security when inspecting. Apart from precarious situations when there may be a fear of falling, there are times when it is difficult to get 'comfortable' on the job. Often the curvature of the aircraft means that whilst one's feet are safely on the platform, the area to which, say, an eddy current probe is to be placed is beyond the normal arm's reach necessitating leaning out of equilibrium. The other arm is then required for support to establish a stable 'working platform' for the hand and probe.

This is fairly obvious but the situation also obtains when the inspector is merely standing doing the job above on a vertical surface. A more stable platform is required than two legs and therefore the other hand is used to lean on. The design of inspection equipment should allow for this. Probes need to be independent of the pressure exerted by the hand wherever possible and test-sets must either be very small and light or be capable of easy safe fixing to the structure or have repeater units for the appropriate parameters which can be read while the probe is being guided.

Can you get at it?.
Above all, if the inspector can't get there, he can't inspect.

Poor access can entail not only difficulty in positioning oneself for an inspection (primary access) but also of seeing once one is there (secondary access). Ideally, the inspector should be able to stand on a level with the structure on job-specific staging. Much more custom-built staging is now available than was seen 10 years ago and much is of good design, although we still see vestiges of the old plank-on-the-oil-barrels-in-the-desert syndrome.

Access to wings is good in general, especially on the underside, although at one installation it was found necessary to reverse the staging because the engine on one side filled the stairs, easy to inpect it's cowling, though.

On top of the wing there is still a tendency to forget that one can step off the edge and there is a definite requirement for non slip mats as the camber can be quite severe and a recumbent inspector or maintenance technician can easily slip off.

Taut wires at hip height along the wing edge could be simply implemented and vortex generator vanes be highlighted with fluorescent tape etc.

Pylon inspection has been seen to be difficult along a wing top and involved the inspector in an uncomfortable and apprehensive few moments lying face down, head below feet, on a slippery surface while reaching around some structure to feel whether a rubber gaiter was intact. The whole area was saturated in repellent (aptly named) and a direct visual inspection was not possible from this angle. This inspection, a short but important one, warranted a cherry-picker but the effort to fetch and position these in some hangars, crossing numerous trailing leads and moving toolboxes (often 2mx2mx1m) is just too much and the less comfortable and therefore less reliable, prone access position may be taken up.

The real problem, seen on a Boeing 707 in the last case, is a generic one for all older aircraft in that it is certainly not economically feasible to make special staging for such small fleet numbers but the extra structural inspections required on aging aircraft make this an urgent consideration.

In one case a crown was inspected from a series of semicircular access bridges from one side to the other. This was admirable for inspecting circumferential production breaks but less handy for working along the crown, which would be required eg in many multiple eddy current tasks.

On many occasions, NDT personnel were seen struggling to keep apparatus in line-of-sight with one hand and operating a probe with the other while leaning forwards uncomfortably due to the fuselage curvature. Simple curved ladder-racks resting on the staging and against the hull would have cured this.

In one hangar two access platforms were joined by a 30ft bridge which flexed in use. Access was by unsecured ladder and then by ducking under the bridge railing (10ft in the air and no hand holds. Due to inadequate maintenance the platform wheels could not be locked and welded angle-iron wedges were used. These still allowed over 1" of travel by the wheels and the castor action allowed their escape from the wedges.

The high tech solution of having a telescopically mounted platform on an overhead gantry as seen in one hangar is probably beyond the means of operators in today's financial climate. One operator, at least, achieves access by dangling inspectors on a running wire from the hangar ceiling.

Steps, mobile staircases and ladders vary enormously in quality and safety. Most have wide bases to avoid tipping and many have hand rails but there are still many that tip easily, that are rickety with loose joints and that have wheels which do not lock. One otherwise sturdy staircase had only one wheel that was lockable and so one moved gradually in a circle during inspection; others could not be adjusted for foot height and rocked continually.

Probably the most dangerous case seen involved steps that were ten feet tall with a top barely large enough for two feet (human that is) so that the scheduled inspection of the forward service door, a complicated enough task involving much torso movement to enable a close scrutiny of a complicated structure, necessitated one to have one foot on the steps and the other on the aircraft: and the next one in the grave?

The maintenance of ladders, steps and scaffolding needs far tighter controls and some form of regular inspection should be made: perhaps even a log book kept for the larger sets. It might also be wise to obtain a few spare wheels etc in case the makers go out of business.

Secondary access Poor secondary access (what you can see when you get there) is more difficult to improve. It is caused by insufficient access panels, lack of headroom (both in height and distance of the eye from the structure to be observed) and pure inaccessibility due to size.

If secondary access is not incorporated in the original design, and the ergonomists have plenty of lists of

normal body lengths, then the only recourse is to mirrors and endoscopes. A small hand held TV camera could aid many inspections where close examination for small cracks is not called although the magnification would have to be large to get the required resolution and this could make for poor sense of location.
Poor secondary access can also be caused by poor specification in the schedule. During one inspection to inspect clevis joint bearing holes on a flap actuating assembly, the schedule failed to take into account the fact that the actuators were still in place; very frustrating.

The ideal height of a platform to inspect and to do the subsequent maintenance are not the same. An inspection is best done at eye-height or a little below whereas maintenance is often best carried out at waist level.

A major improvement to access is seen where trailing services such as electrical cables and air hoses are brought from a central point or line underneath the aircraft. Moving steps and staging is quicker and safer. The cost of installing such a system is small compared with a fatality and much time is saved which is now spent on repairing broken leads, and air-lines.

What can we do?
It is doubtful whether it will be possible in the foreseable future to provide quantitative assessments.
In this industry, the provision of a complete Task Analysis, ie a breakdown of each task and the parameters affecting it, the presentation of the problem, the basic knowledge required by the inspector and the probabilities of success in searching, finding and assessing the problems would be a vast task. It is debatable whether a task analysis covering any one aircraft is possible within its own lifetime.

The research necessary to produce even the simplest quantitative assessment of, say, the effect of lighting on visual scanning would be lengthy and could not include the effects of other every other significantly interacting parameter such as operator comfort, eyesight, and noise etc. The full assessment of all these or any other parametric scheme has to remain a pipe-dream. We can only ensure that there is plenty of light available.

In the end what we must do is accept that there are factors, human and mechanical, which are unquantifiable but possibly significant and eliminate them or at least minimise their effect. This can be done by ensuring that the operator is considered above all else in such areas as comfort on the job, training, information feedback, working conditions and, most of all respect.

Reliability, as a discipline is inherently difficult to quantify as mostly we are dealing with statistically insignificant factors. Fortunately, aircraft engineering is routed in good engineering practice and its workers have a genuine love of the end product. It is this fact that ensures above all, a safe aircraft.

Acknowledgements

Although precluded by contract from thanking here personally any particular operations personnel for their real help, they know who they are and I do. Also Prof Colin Drury at SUNY Buffalo for his enthusiasm and hospitality while I was in the USA and the staff of the CAA and FAA who opened so many doors.

REFERENCES

1. Lock MWB & Strutt JE **"Reliability of maintenance of transport aircraft"** CAA report 85013: ISBN 0 86039 251 1

2. Thackray RI **"HF evaluation of the work environment of operators engaged in the inspection and maintenance of aging aircraft"** DOT/FAA/AM-92/3 Office of Aviation medicine Washington DC 20591

3. Drury CG & Lock MWB **"Reliability in Aircraft Inspection: UK and USA perspectives"** CAA paper 94001, ISBN 0 86039 567 7

4. Lock MWB **"Aspects of reliability of aircraft structure NDT inspection"** Insight v36:10, p748-, Oct 1994

5. Lock MWB **"Human and workplace factors affecting reliability in aircraft inspection** Insight v37:2, p83-, Feb 1995

6. Spencer WS & Drury CG **"Visual inspection research project report on benchmark inspections"** DOT/FAA/AR-96/65

7. Drury CG (1988)**'The Human Operator as an Inspector'**; Report of FAA meeting: Human Factors Issues in Aircraft Maintenance and Inspection: Oct 1988 Virginia.

8. Hockey (1978)**'Effects of Noise on Human Work Efficiency'**; 'Handbook of Noise Assessment' chapter 14, pub. van Nostrand Reinhold.

PRACTICAL EVALUATION OF CRACK DETECTION CAPABILITY FOR VISUAL INSPECTION IN JAPAN

H. Asada and T. Sotozaki
Airframe Division, National Aerospace Laboratory
6-13-1 Osawa, Mitaka, Tokyo 181-0015, Japan

S. Endoh and T. Tomita
Airworthiness Division, Civil Aviation Bureau
Ministry of Transport
2-1-3 Kasumigaseki, Chiyoda, Tokyo 100-8989, Japan

SUMMARY

The role of visual inspection is important for maintaining and improving aircraft structural integrity. The Civil Aviation Bureau of Japan organized the investigation team for visual inspection capability consisting of three major operators, four major manufacturers and the National Aerospace Laboratory. This paper describes the collected field data of cracks detected by visual inspection during maintenance of aircraft operated by Japanese airlines, the analyzed results and the significant information for safety and reliability of aircraft structures evaluated by the damage tolerance design. Detected cracks are collected from primary aluminum alloy structures of in-service transport aircraft. The number of detected cracks is more than 1000 collected over a period of three years.

1. INTRODUCTION

The damage tolerance design for transport aircraft structures requires that damages should be detected and repaired before reaching critical size under adequately planned maintenance inspection programs. In consequence, it is needless to say that the role of operators is augmented and the capability of inspectors becomes significant for maintaining aircraft structural safety designed by the damage tolerance regulations.

Inspection programs are generally developed by manufacturers with operators' support. Operators apply thus proposed inspection programs to their in-service aircraft for detecting timely damages before they become critical sizes.

Although a fundamental structural inspection is conducted by visual inspection, non-destructive inspection methods are pertinently employed for structural element inspection. These inspection methods have been improved and new methods have been developed in order to revise detection capability of damage. However, it should be noted that the role of visual inspection would not diminish from the view point of structural safety hereafter. In spite of importance of visual inspection, data which can evaluate its capability have not been collected in Japan.

As a result of the investigation on the aircraft accident of Japan Airlines' Boeing 747 SR-100, JA 8119, occurred on August 12, 1985, the Aircraft Accident Investigation Commission of Japan (JAAIC) made a proposal on collection and analysis of visual inspection data to the Minister of Transport, dated on June 19, 1987. This proposal demands that a study should be made with respect to discovery of cracks by visual inspection for the improvement of aircraft maintenance technology. In addition, it mentions as follows:

- In most cases, discovery of cracks caused on aircraft structures has been made by visual inspection. However, no sufficient reference is presently available on the problem to determine to what extent the visual inspection is effective in discovery of cracks.

- It is necessary to study measures to improve aircraft maintenance technology by collection and analysis of data on crack discovery by visual inspection on transport aircraft in current use in Japan.

The target of this proposal was to point out the significance of visual inspection capability.

The Airworthiness Division of the Civil Aviation Bureau of Japan (JCAB) organized the investigation team in September, 1987, and then the team immediately started its activity for collecting field data of cracks detected by visual inspection in order to analyze the actual circumstance for visual inspection capability in Japanese airlines. The members of the team consisted of the JCAB, three major operators which were Japan Airlines, All Nippon Airways, Japan Air System, four major aircraft/engine manufacturers which were Mitsubishi, Kawasaki, Fuji and Ishikawajima-Harima Heavy Industries, the Association of Air Transport Engineering and Research, and the National Aerospace

Paper presented at the RTO AVT Workshop on "Airframe Inspection Reliability under Field/Depot Conditions", held in Brussels, Belgium, 13-14 May 1998, and published in RTO MP-10.

Laboratory. During this investigation, the accident of Aloha Airlines, Boeing 737-200, occurred on April 28, 1988. This accident emphasized the urgent re-evaluation for continuing airworthiness of aging aircraft structures again.

This report describes the collected field data of cracks detected by visual inspection, the analyzed results and the important information of these data(Ref 1). The investigation on visual inspection capability in Japan was carried out for about three years between January in 1988 and November in 1990. The data of detected cracks more than 1000 were collected from primary aluminum alloy structures. Several detected crack surfaces were fractographically examined to investigate their fatigue crack propagations.

2. PURPOSE AND PROCEDURE OF INVESTIGATION

This investigation consists of following four purposes:

(1) To collect data of cracks detected by visual inspection implemented for in-service transport aircraft operated by Japanese airlines.

- Following three methods were discussed for this investigation:

  1) To apply field data collected during maintenance inspection.
  2) To perform round robin tests using structural elements with cracks which are removed from inspected in-service structures(Ref 2).
  3) To perform simulation tests which use developed standard elements with cracks.

Finally, the field data method were accepted, because it was concluded that the round robin test and the simulation test methods were difficult to reappear pertinently actual field circumstances of visual inspection.

(2) To analyze crack detection capability of visual inspection in Japan.

(3) To provide data bases on information, knowledge and analytical results of cracks detected by visual inspection.

(4) To study application of the data bases.

3. DATA SHEET FORMAT

As crack data are collected and reported during maintenance by operators' inspectors, a data sheet should be simple and clear for inspectors to check each item defined precisely in order to reduce inspectors' burden. It is agreed that a fatigue crack detected in a primary structural element made of aluminum alloys should be reported on the data sheet. A crack caused by corrosion or stress corrosion is scarcely included in this investigation.

The data sheet mainly consists of following three items, namely, (1) General information, (2) Inspection detail, and (3) Crack data as shown in **Figure 1**. These items with related subitems and contents are explained below.

(1) General information

A. Aircraft type model.
Ten types operated by Japanese airlines are accepted on this data sheet. Therefore, new versions belong into their prototypes respectively. For example, MD-80 is classified as DC-9.

B. Inspection date: month / year.
The date when a crack is detected is reported. In addition, the numbers of flight hours and flights accumulated by the aircraft are also acquired.

C. Maintenance activity.
In order to investigate the relation between inspection level and crack detection, the maintenance activity detecting a crack is reported. "Others" indicates maintenance activity corresponding with D-check and H-maintenance, which is higher than C-check.

D. Structural area.
Aircraft structures are divided into five areas.

E. External or internal location.
Locations of detected cracks are classified by external or internal location for identifying that cracks emanating on external skin are easily detected.

(2) Inspection detail

A. Inspection type
Inspection types are classified as follows:

- Engineering order: An inspection performed according to a service bulletin.
- Special inspection: A supplemental or an additional inspection, or a special work order planned by operators.
- Task card at regular maintenance: A zonal inspection and a significant structural inspection specifying a zone and a structure element to be inspected, respectively.

•Others: Inspections except for the types mentioned above.

B. Prior information
This information means that an inspector recognizes in advance that a crack may emanate in an inspected structural element. It is mentioned that this is the most significant factor for detecting a crack.

C. Inspection distance
This distance is divided into three ranges which are accepted to be significant for detecting a crack in a structural element.

D. Surface treatment
The coating condition of an inspected structural element is reported such as primer for corrosion prevention or top coat painted on primer.

E. Surface condition
The surface condition is reported as dirty or clean, which is considered to be important for crack detection.

(3) Crack data

A. Visible crack length
The uncovered crack length is reported on this sheet. The crack length that an inspector initially and visually measures is reported as the initial apparent crack length. After that, by using visual aids or NDI, the inspector accurately measures its length which is reported as the actual crack length. In case of multiple cracks which are frequently detected in an inspected structural element, the longest crack length is reported.

B. Open or closed crack
This subitem reports that a detected crack is opening or closed. This is considered to have an effect on the crack detection capability.

C. Crack origin
Origins are classified into three parts which are fastener hole, edge and others which are closely related to areas where inspectors pay their attention to detect cracks.

D. Leak indication
It is also expected that this subitem is effective to an inspector's crack detection.

4. COLLECTED CRACK DATA

Cracks were collected between January in 1988 and November in 1990 by three Japanese operators, Japan Airlines, All Nippon Airways and Japan Air System. The total number of detected cracks is 1054 which is summarized in **Table 1**. The number of cracks detected from B747s accounts for about 68% of the entire data. The numbers of total aircraft and B747s with detected cracks are 159 and 61, respectively. Several aircraft were inspected twice or three times during this investigation period. Values in parentheses indicate percentages. The result of each subitem is discussed below.

A. Maintenance activity
The majority of cracks are detected during C-check and "C/Others" for B747. On the other hand, cracks in other aircraft are detected during C-check. "C/Others" denotes that cracks are detected at the maintenance activity when C-check and "Others" are implemented at the same time.

B. Structural area
A large number of cracks are detected in fuselage structural elements, and especially as for B747, more than 90% of cracks are detected in fuselage structural elements.

C. External or internal location
Almost all cracks are detected in internal structural elements. Cracks are scarcely detected from external surfaces.

D. Inspection type
Many cracks are detected during inspections performed by engineering order for B747. In case of other aircraft, more than 50% of cracks are detected by zonal inspections of task card.

E. Prior information
Almost all cracks are detected by inspectors with prior information.

F. Inspection distance
Almost all cracks are detected under the inspection distance of less than 50cm.

G. Surface treatment
A large number of cracks are detected from the surface treatment of primer.

H. Surface condition
Almost all cracks are detected on clean condition.

I. Single crack or multiple cracks
The majority of detected cracks are single cracks. An precise crack length is visually confirmed.

J. Open or closed crack
The number of closed cracks is almost twice that of detected open cracks.

K. Crack origin
Although it is generally considered that most of

cracks emanate from fastener holes, the present result shows that more cracks emanate from edges than from fastener holes.

L. Leak indication

Almost all cracks in fuselage structural elements are detected without leak indication such as tar.

M. Others

The shortest length of the reported cracks is about 0.02 inches(0.5mm) detected in a fuselage structural element of B737 at D-check with prior information, and the longest one is about 14.5 inches(368cm) which is a single crack detected in a fuselage structural element of B747 at C-check without prior information.

5. ANALYTICAL RESULT OF COLLECTED CRACK DATA

Figure 2 shows the monthly frequency distributions of detected cracks for B747, other aircraft denoted as "Others" and "All" which is the sum of B747 and "Others" with or without inspectors' prior information. The monthly frequencies of B747 increased when the aircraft with many cracks were inspected. **Figure 3** indicates the relation between number of flights and frequency of detected cracks. It is recognized that the majority of cracks for B747 are detected between 15,000 and 25,000 flights. As for "Others" including 9 aircraft type models, the remarkable tendency can not be found in this relation. **Figure 4** depicts the frequency distributions of detected crack length of B747, "Others" and "All". The length of many cracks detected in B747 and "Others" is shorter than 1 inch(25.4mm).

The relations of the subitems between visible crack length and normalized cumulative relative frequency are indicated in **Figures 5** to **16**. The normalized cumulative relative frequency is given by a value of a cumulative frequency divided by a total number shown in parentheses of each figure. The discussion of the influences of the subitems for crack detection are given below.

A. Prior information

Figure 5 shows the influence of prior information on detected crack length. The results indicate that the detected crack length with prior information becomes shorter than that without prior information. As for "All", the percentage of detected crack length shorter than 1 inch is about 75% in the case with prior information. On the other hand, the percentage reduces to about 50% in the case without prior information. The percentage of detected crack length of B747 with or without prior information is slightly more than that of "Others".

The comparisons of detected crack length between the Federal Aviation Administration(FAA) data (Dinkeloo and Moran(Ref 3), Goranson and Hall(Ref 4), and Goranson and Rogers(Ref 5)) and the present results are shown in **Figure 6**. The FAA results are derived by using the data reported in the Mechanical Reliability Report(MRR) and the Service Difficulty Report(SDR) submitted from 1963 through 1973. The data are classified into cracks detected by visual inspection, non-destructive inspection(NDI) and "All" which is the sum of visual inspection and NDI. The FAA results with prior inspection correspond to directed inspection with service bulletin or airworthiness directive. The results without prior information show non-directed or general area inspection without service bulletin or airworthiness directive. In the case of non-directed inspection, almost all cracks are detected by visual inspection. As for directed inspection, the majority of cracks are detected by NDI. The FAA results and the present results are shown by stepped solid lines and dotted lines, respectively.

Figure 6.1 depicts the results without prior information. The present results of visual inspection at a certain length are superior to those of FAA at the same length. The tendency is remarkable for the result of B747. The results with prior information are shown in **Figure 6.2**. The present results of B747 and other aircraft are quite superior to the FAA results of visual inspection. It should be noted to consider that the FAA data were collected between 1963 and 1973, namely, about 20 years before the present investigation performed between 1988 and 1990.

B. Maintenance activity

Figure 7 shows the relation between inspection levels detecting cracks and visible crack length. The activities are classified into C-check and the checks higher than C-check including D-check, "Others" and C/Others-check. The result depicts that inspection levels have no influence on the relation between visible crack length and normalized cumulative relative frequency.

C. Structural area

The influence of structural area on visible crack length is shown in **Figure 8**. It is noticed that remarkable differences among structural areas are not found.

D. External or internal location

Figure 9 shows that locations have no difference of detected crack length. In other words, the result can not clarify that cracks on external surface can easily be detected. It is pointed out that the number of cracks detected on external surface is very few in this investigation.

E. Inspection type

Figure 10 indicates the influence of inspection types. The result is considered to show that the inspection types have no remarkable influence on detected crack length.

F. Inspection distance

The relation between inspection distance and visible crack length in **Figure 11** shows the understandable result that a detected crack length under a shorter inspection distance becomes shorter.

G. Surface treatment

The influence of surface treatment is shown in **Figure 12.** The notable difference can not be found from this result.

H. Surface condition

Figure 13 indicates the influence of surface condition. It is observed that a detected crack length on the dirty surface becomes longer than that on the clean surface except for the range of shorter crack length. It should be noted that the number of cracks detected under the dirty condition is very few.

I. Open or closed crack

According to **Figure 14,** the length of detected closed crack is shorter than that of detected open crack. The fact is different from general recognition that open cracks detected by inspectors are shorter than closed cracks. As it is assumed that shorter cracks are closed and longer cracks are opened, the present result does not indicate that the detection of open crack is worse than that of closed crack.

J. Crack origin

The length of detected crack emanating from fastener hole or edge becomes shorter than that of crack emanating from "Others" as shown in **Figure 15.** Therefore, the detection of crack emanating from "Others" is difficult.

K. Leak indication

Figure 16 depicts that the distribution of detected cracks with leak indication is almost the same as that without leak indication. However, it should be pointed out that the number of detected cracks with leak indication is very few.

6. CONCLUSIONS

The role of visual inspection is significant for maintaining and improving the safety and reliability of aircraft structures evaluated under the damage tolerance principle. Recently, many non-destructive inspection methods have been developed and successfully applied to aircraft structures. However, visual inspection will be widely used for structural maintenance in future.

Over a period of three years, the data of cracks detected by visual inspection were collected during the maintenance of aircraft operated by Japanese airlines. The number of detected cracks amounted to 1054, and the collected data constituted the data on visual inspections in Japan. Effective results which can evaluate visual inspection capability are obtained by analyzing the crack data. The subitems and their related contents on data sheets are enough to evaluate the visual inspection capability of detecting cracks in Japanese airlines.

The number of detected cracks is 1054 which can provide the data base on visual inspection in Japan, and it is expected that further collection and investigation will be performed in succession. The data summarized with the aid of the developed computer program puts emphasis on the following reliable evaluation:

- The great majority, about 90%, of cracks are detected with inspectors' prior information. This result shows that inspectors should adequately be provided with crack information.

- A large number of cracks are detected from structural elements of internal fuselage. The tendency is much remarkable for B747.

- The present data of detected cracks are compared with the result reported by the MRR/SDR submitted to the FAA. Although it is realized that the periods of data collection are different each other, the result on visual inspection capability of Japanese airlines is equal or superior to that of FAA. It is recognized that visual inspection of detecting a crack successfully plays an important role for ensuring the structural safety and integrity of in-service transport aircraft operated in Japan.

Based on the results explained in the previous chapter, the subitems accepted in this investigation are classified into factors with or without influence on detected crack length as follows:

(1) Factors with influence
- Prior information
- Inspection distance
- Surface condition
- Crack origin

(2) Factors without influence
- Maintenance activity
- Structural area
- Inspection type

The influence of the following factors on detected crack length can not be evaluated from the present

investigation, because the numbers of the data of those factors are not enough to reach the definite conclusion.

- External or internal location
- Surface treatment
- Open or closed crack
- Leak indication

ACKNOWLEDGMENTS

The authors are deeply indebted to Dr. S. Sampath of the Federal Aviation Administration and Mr. D. Simpson of the National Research Council who recommended us to present this paper in this Workshop. The authors also thank Japan Airlines, All Nippon Airways, Japan Air Systems, the Association of Air Transport Engineering and Research, Mitsubishi, Kawasaki, Fuji and Ishikawajima-Harima Heavy Industries, and Dr. S. Ito of the National Aerospace Laboratory for their remarkable contribution and valuable discussion of this investigation.

REFERENCES

1. Endoh,S., Tomita,H., Asada,H. and Sotozaki,T., "Practical Evaluation of Crack Detection Capability for Visual Inspection in Japan", Proceedings of the 17th Symposium of the International Committee on Aeronautical Fatigue(ICAF), 1993, pp.259-280.
2. Fahr, A., Forsyth, D. & Wallace, W., "POD Assessment of NDI Procedures Using a Round Robin Test", AGARD-R-809, January 1995.
3. Dinkeloo,C.J. and Moran,M.S., "Structural Area Inspection Frequency Evaluation (SAIFE)", Federal Aviation Administration, U.S. Department of Transportation, Report No. FAA-RD-78-29, 1978.
4. Goranson,U.G. & Hall,J., The Aeronautical Journal, Vol.84, No.838, 1980, pp.374-385.
5. Goranson,U.G. & Rogers,J.T., "Elements of Damage Tolerance Verification", The 12th Symposium of the International Committee on Aeronautical Fatigue(ICAF), 1983.

Table 1 Summary of detected crack data

Item	Subitem	Content	All	B747	Others
1. General information	Number of cracks		1054(100)	714(67.7)	340(32.3)
	Number of aircraft		159(100)	61(38.4)	98(61.6)
	Maintenance activity	Line	4(0.4)	3(0.4)	1(0.3)
		C-check	463(43.9)	259(36.3)	204(60.0)
		D-check	77(7.3)	- (-)	77(22.7)
		Others	221(21.0)	175(24.5)	46(13.5)
		C/others	289(27.4)	277(38.8)	12(3.5)
		Total	1054(100)	714(100)	340(100)
	Structural area	Fuselage	788(74.8)	651(91.3)	137(40.3)
		Wing	113(10.7)	31(4.4)	82(24.1)
		Empennage	44(4.2)	1(0.1)	43(12.6)
		Pylon	48(4.6)	25(3.5)	23(6.8)
		Door	60(5.7)	5(0.7)	55(16.2)
		Total	1053(100)	713(100)	340(100)
	External or internal loc.	External surface	75(7.2)	32(4.5)	43(12.8)
		Others	972(92.8)	678(95.5)	294(87.2)
		Total	1047(100)	710(100)	337(100)
2. Inspection detail	Inspection type	Engrg. order	448(43.4)	395(55.5)	53(16.5)
		Special	56(5.4)	20(2.8)	36(11.2)
		Task card			
		:Zonal	264(25.6)	76(10.7)	188(58.6)
		:Significant	214(20.7)	194(27.2)	20(6.2)
		Others	51(4.9)	27(3.8)	24(7.5)
		Total	1033(100)	712(100)	321(100)
	Prior information	With	923(87.6)	632(88.5)	291(85.6)
		Without	131(12.4)	82(11.5)	49(14.4)
		Total	1054(100)	714(100)	340(100)
	Inspection distance	50cm >	989(93.8)	680(95.2)	309(90.9)
		50cm~1m	59(5.6)	32(4.5)	27(7.9)
		1m <	6(0.6)	2(0.3)	4(1.2)
		Total	1054(100)	714(100)	340(100)
	Surface treatment	Primer	775(73.5)	565(79.1)	209(61.6)
		Top coat	242(23.0)	134(18.8)	108(31.9)
		Not coated	37(3.5)	15(2.1)	22(6.5)
		Total	1054(100)	714(100)	339(100)
	Surface condition	Dirty	48(4.6)	28(3.9)	20(6.0)
		Clean	997(95.4)	682(96.1)	315(94.0)
		Total	1045(100)	710(100)	335(100)
3. Crack data	Single crack or multiple cracks	Single	859(81.8)	589(82.5)	270(80.4)
		Multiple	191(18.2)	125(17.5)	66(19.6)
		Total	1050(100)	714(100)	336(100)
	Measurement of crack length	Visual	727(71.3)	531(76.1)	196(60.9)
		NDI	293(28.7)	167(23.9)	126(31.9)
		Total	1020(100)	698(100)	322(100)
	Open or close crack	Open	374(35.6)	224(31.5)	150(44.1)
		Closed	676(64.4)	486(68.5)	190(55.9)
		Total	1050(100)	710(100)	340(100)
	Crack Origin	Fastener hole	228(21.7)	136(19.1)	92(27.2)
		Edge	532(50.5)	397(55.6)	135(39.8)
		Others	167(15.9)	95(13.3)	72(21.2)
		Fast. hole & edge	104(9.9)	77(10.8)	27(8.0)
		Fast. hole & others	10(0.9)	3(0.4)	7(2.1)
		Edge & others	12(1.1)	6(0.8)	6(1.8)
		Total	1053(100)	714(100)	339(100)
	Leak indication (Tar or others)	With	61(5.8)	38(5.3)	23(6.8)
		Without	990(94.2)	674(94.7)	316(93.2)
		Total	1051(100)	712(100)	339(100)

Each total number of all aircraft data is not always equal to 1054 because of incompletely reported data.

· (·) : %

1. GENERAL INFORMATION

A. AIRCRAFT TYPE MODEL
☐B-747 ☐B-767 ☐B-727 ☐B-737 ☐DC-10 ☐DC-9 ☐L-1011 ☐A-300 ☐A-320 ☐YS-11

B. INSPECTION DATE _____ 19____
MONTH YEAR

C. MAINTENANCE ACTIVITY
☐ LINE MAINTENANCE
☐ C-CHECK MAINTENANCE
☐ D-CHECK MAINTENANCE
☐ OTHERS (HIGHER THAN C-CHECK)

D. STRUCTURAL AREA
☐ FUSELAGE
☐ WING
☐ EMPENNAGE
☐ PYLON
☐ DOOR(INCLUDING LANDING GEAR DOOR)

E. EXTERNAL OR INTERNAL LOCATION
☐ EXTERNAL SURFACE
☐ OTHERS

2. INSPECTION DETAIL

A. INSPECTION TYPE
☐ ENGINEERING ORDER
☐ SPECIAL INSPECTION
☐ TASK CARD AT REGULAR MAINTENANCE — ☐ ZONAL INSPECTION / ☐ SIGNIFICANT STRUCTURAL INSPECTION
☐ OTHERS

B. PRIOR INFORMATION
☐ WITH
☐ WITHOUT
HAVE YOU HAD ANY PRIOR INFORMATION FOR POSSIBLE CRACKS IN THE AREA ?

C. INSPECTION DISTANCE
☐ LESS THAN 50cm
☐ FROM 50cm TO 1m
☐ 1m AND UP

D. SURFACE TREATMENT
☐ PRIMER
☐ TOP COAT
☐ NOT COATED

E. SURFACE CONDITION
☐ DIRTY
☐ CLEAN

3. CRACK DATA

A. VISIBLE CRACK LENGTH
a INITIAL APPARENT CRACK LENGTH ___ INCH — ☐ SINGLE CRACK / ☐ MULTIPLE CRACKS
b ACTUAL CRACK LENGTH ___ INCH INSPECTION METHOD — ☐ VISUAL / ☐ NDI (INCLUDING DYE PENETRANT INSPECTION

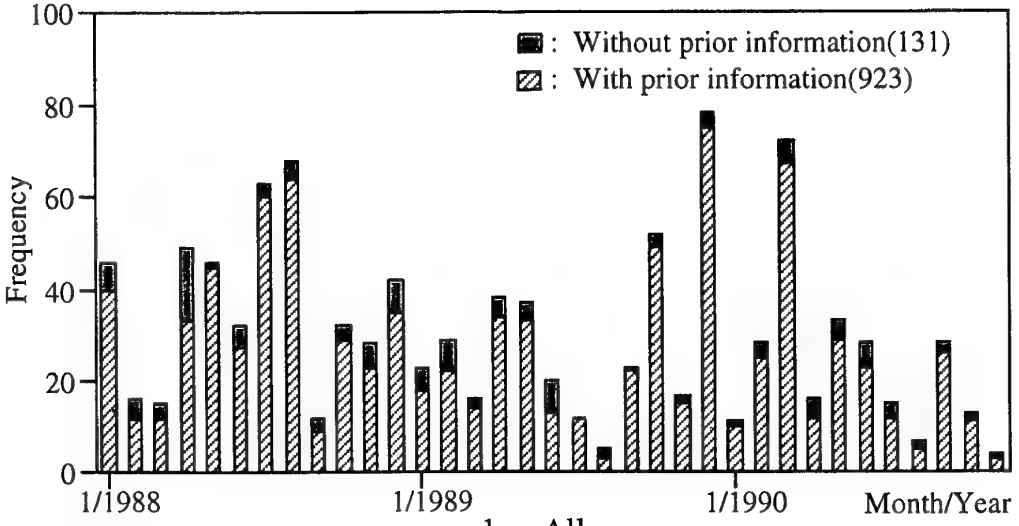
B. OPEN OR CLOSED CRACK
☐ OPEN CRACK (INCLUDING FRACTURE)
☐ CLOSED CRACK

C. CRACK ORIGIN
☐ FASTENER HOLE
☐ EDGE
☐ OTHERS
} (CHECK TWO ORIGINS FIR ONE CRACK IF NECESSARY.)

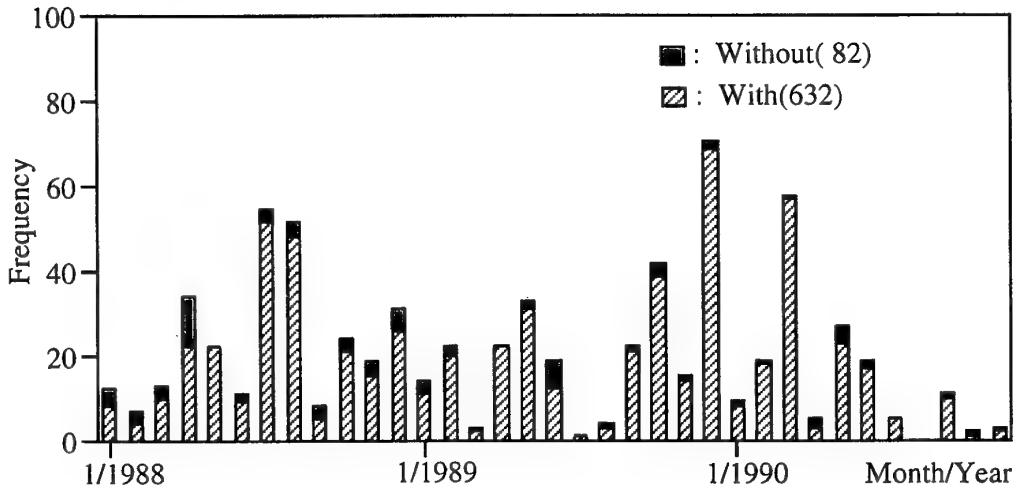
D. LEAK INDICATION (TAR OR OTHERS)
☐ WITH
☐ WITHOUT

* 1 inch = 25.4mm

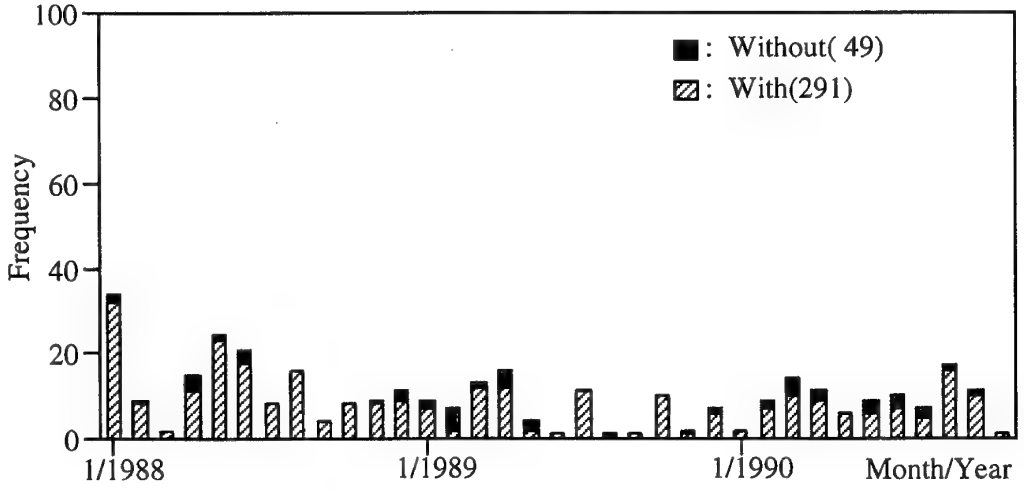
Figure 1 Crack data sheet for visual inspection



1. All



2. B747



3. Others

Figure 2 Monthly frequency distribution of detected cracks

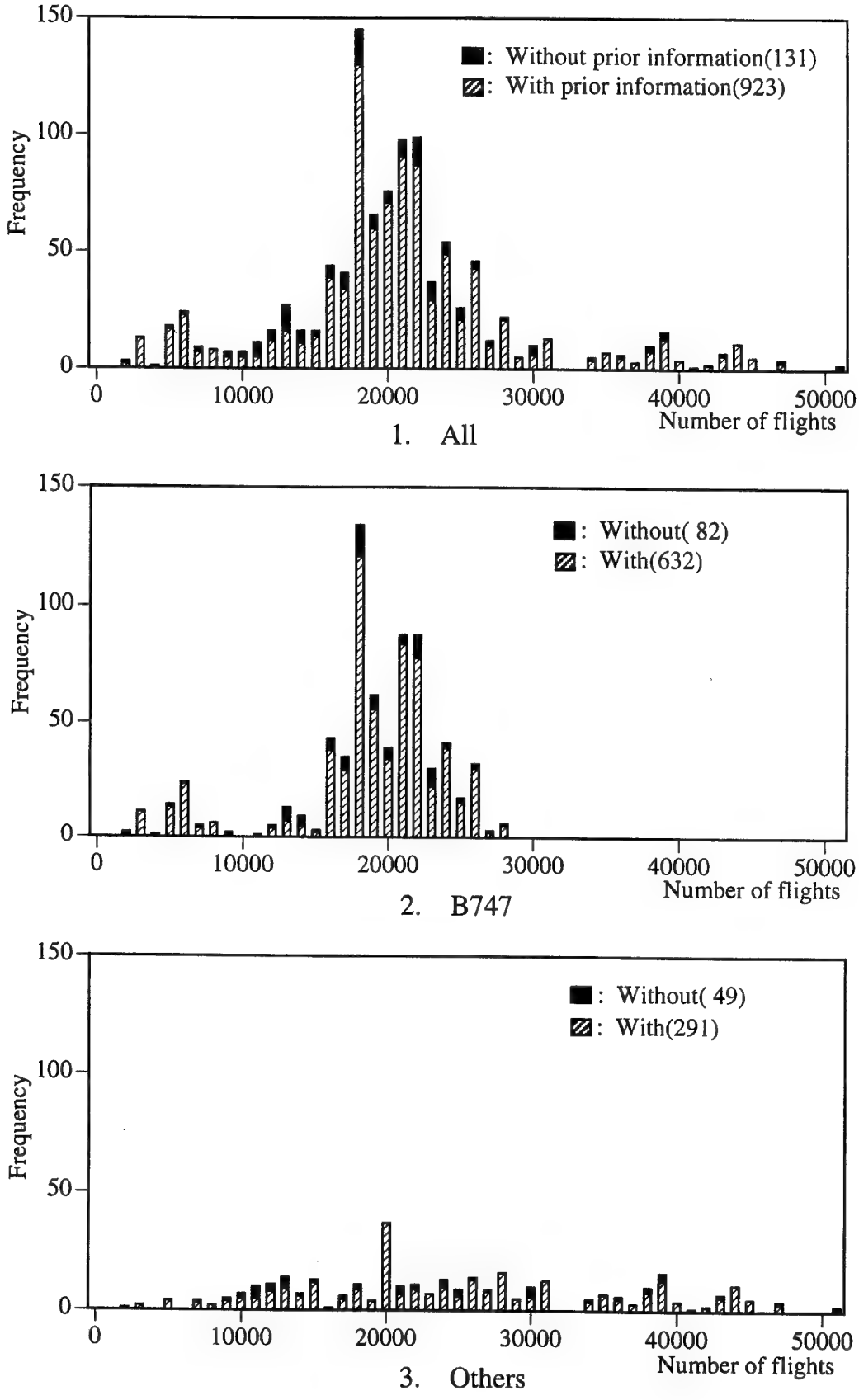


Figure 3 Relation between number of flights and frequency of detected cracks

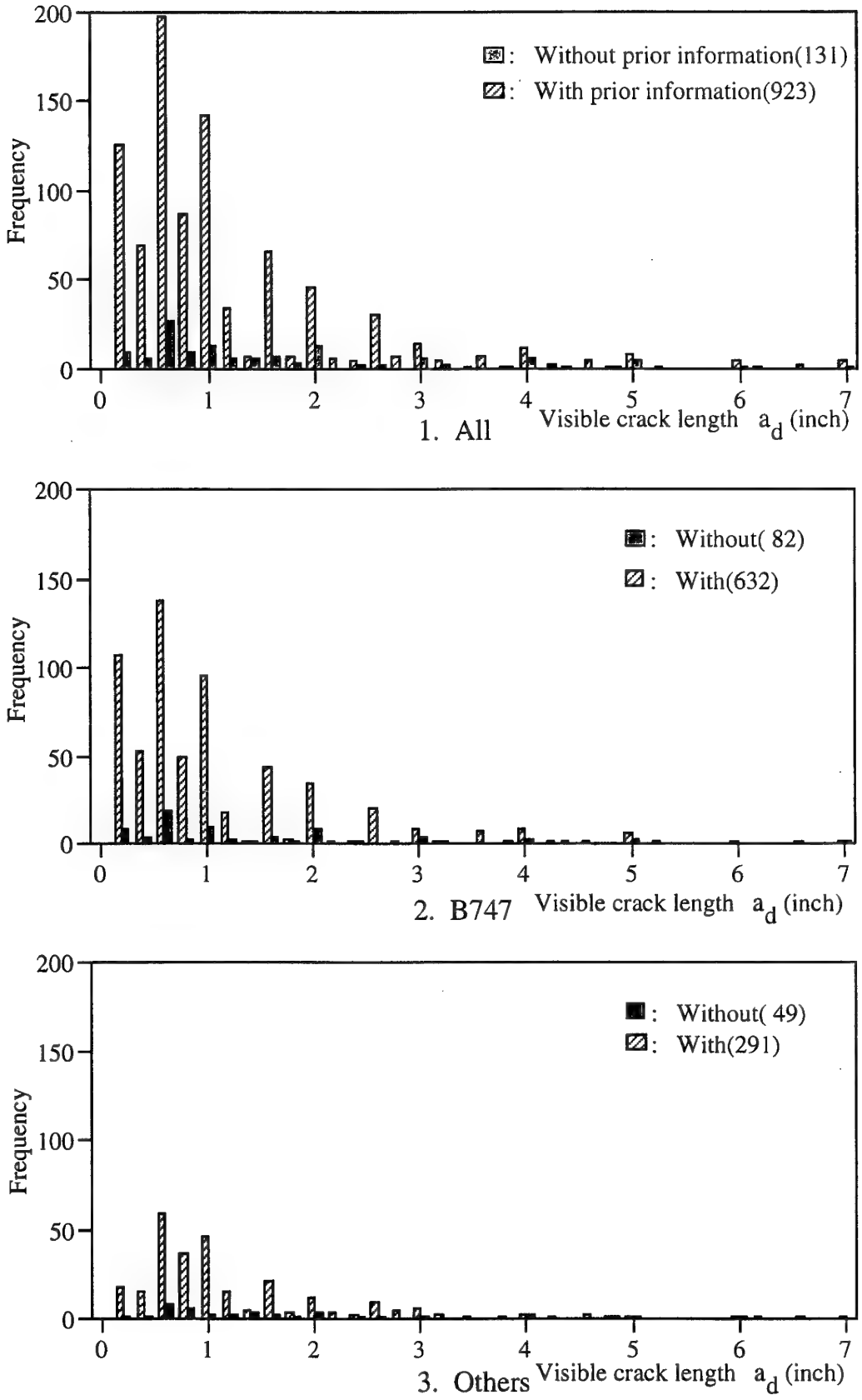


Figure 4 Frequency distribution of detected crack length

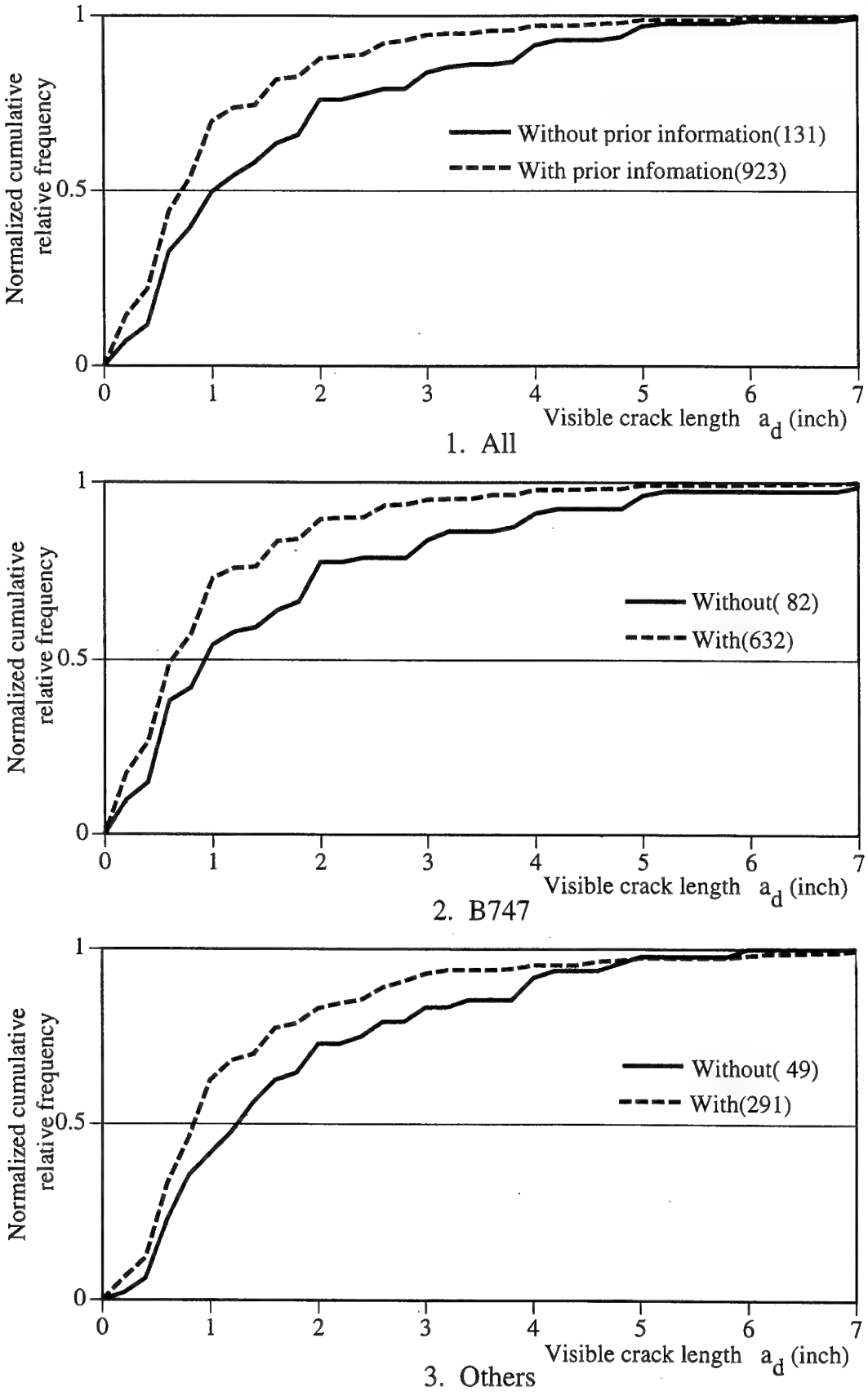
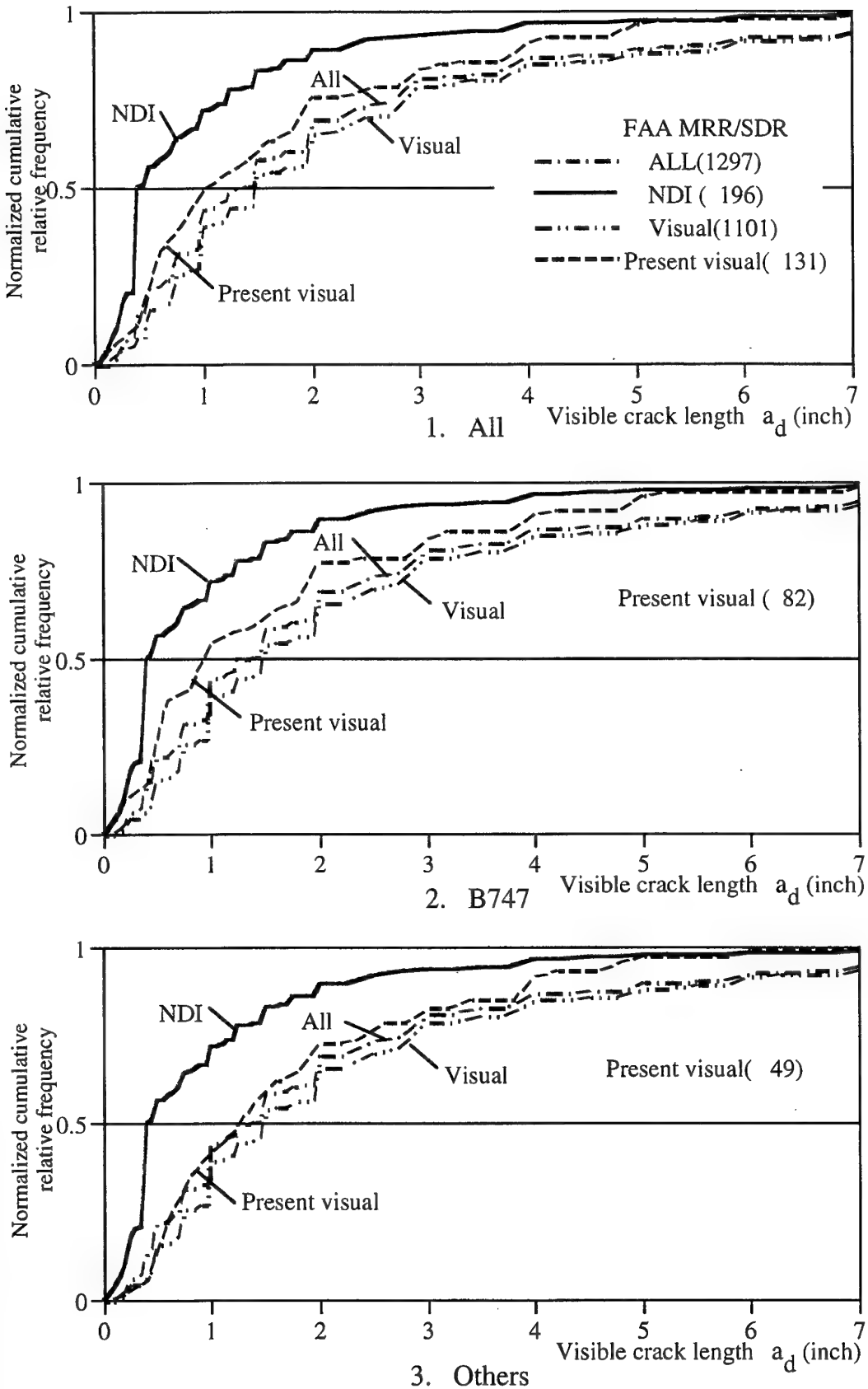
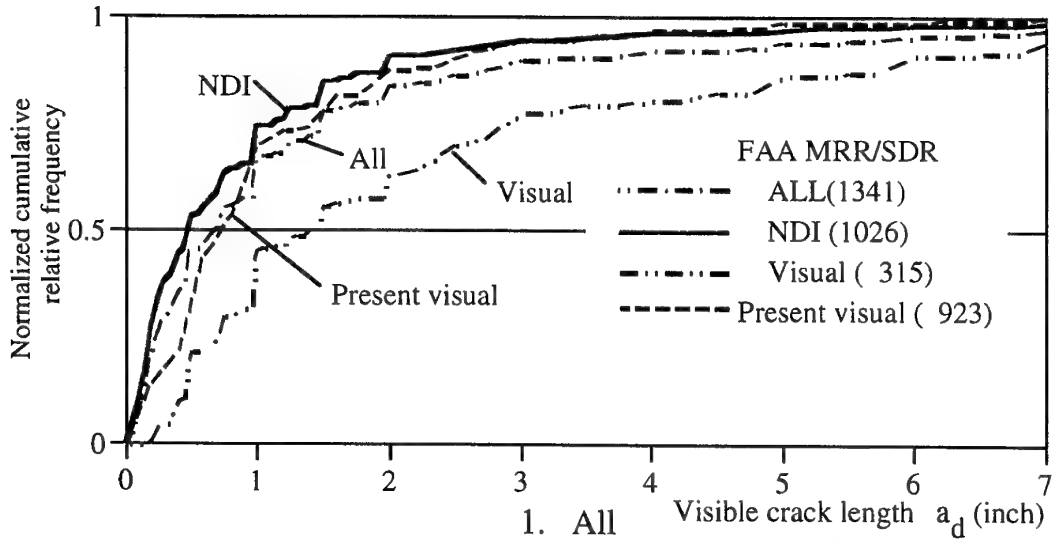


Figure 5 Influence of prior information on detected crack length

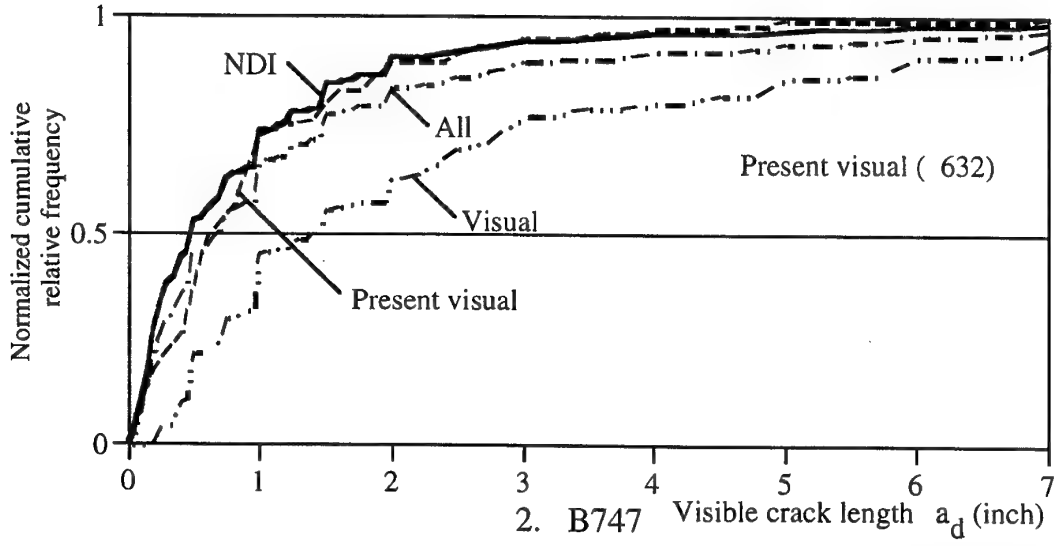


1. Crack detected without prior information

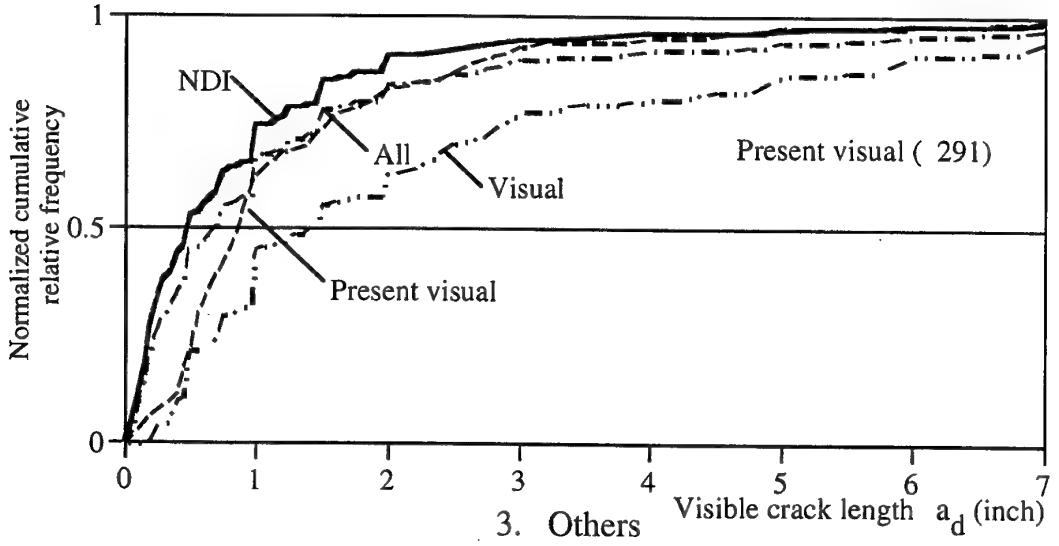
Figure 6 Comparison of FAA and present results for detected crack length



1. All



2. B747



3. Others

2. Crack detected with prior information

Figure 6 Comparison of FAA and present results for detected crack length (Continued)

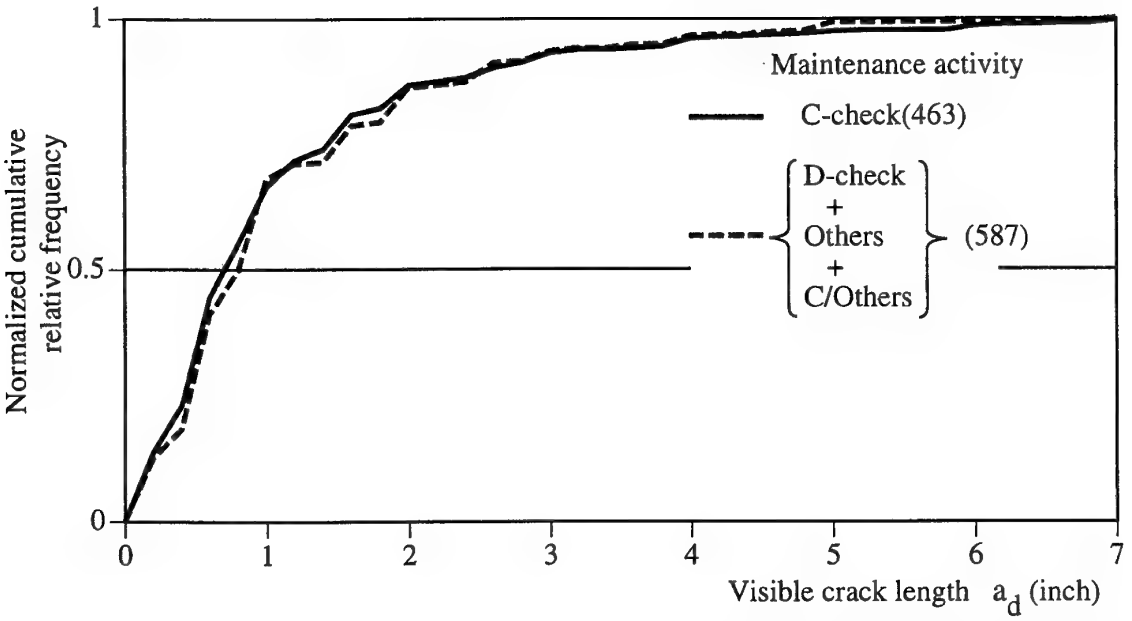


Figure 7 Influence of maintenance activity on detected crack length

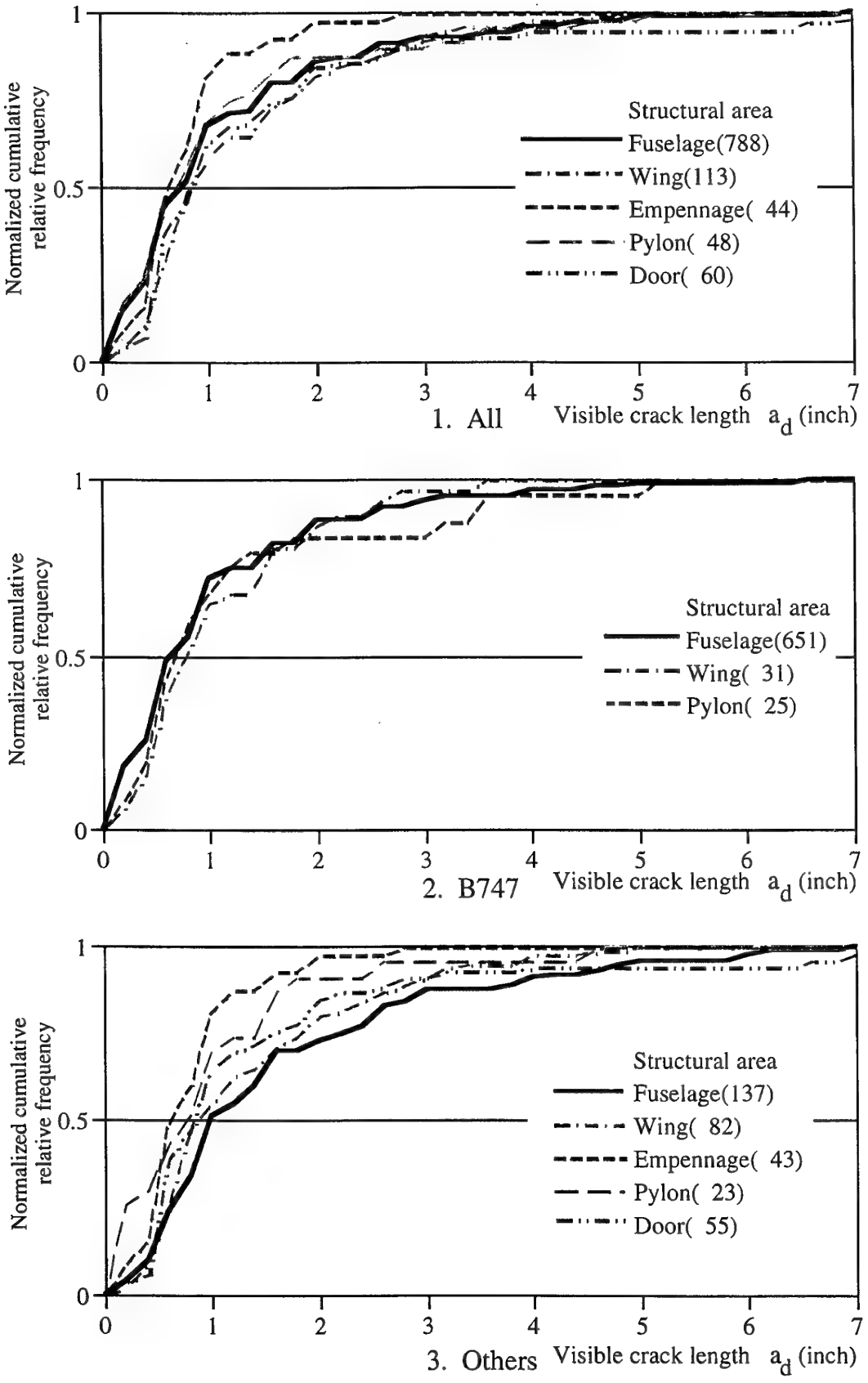


Figure 8 Influence of structural area on detected crack length

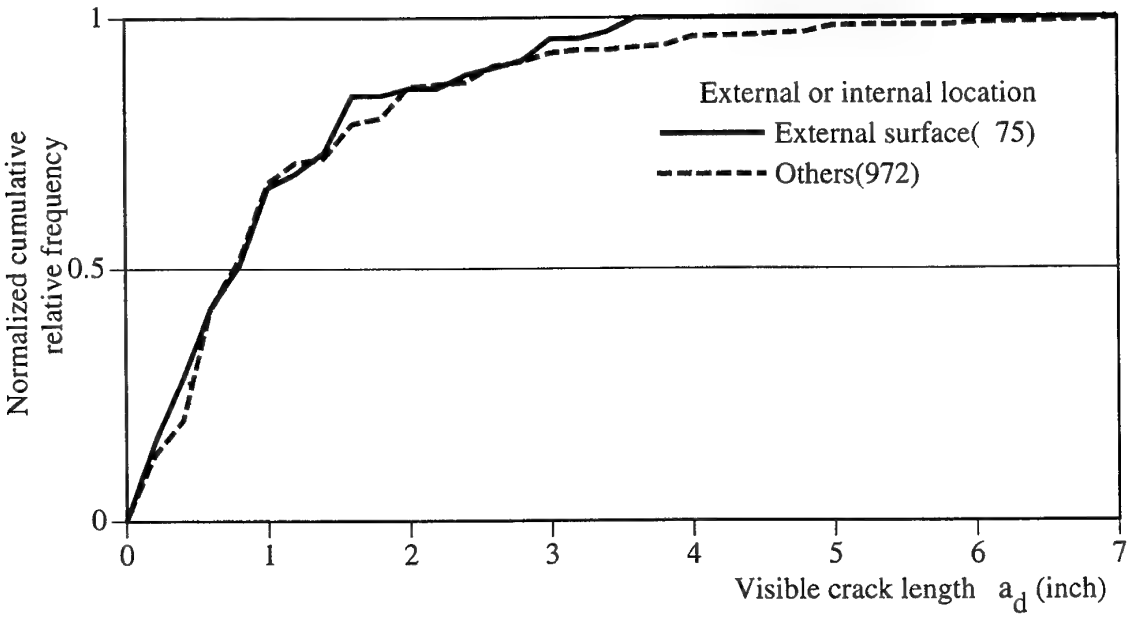


Figure 9 Influence of external or internal location on detected crack length

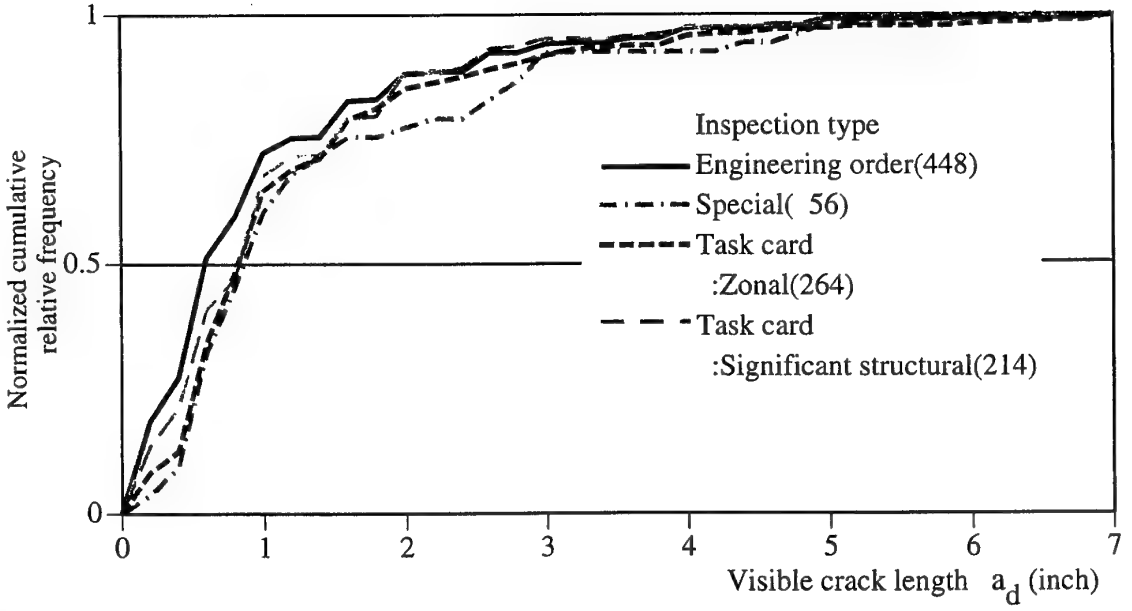


Figure 10 Influence of inspection type on detected crack length

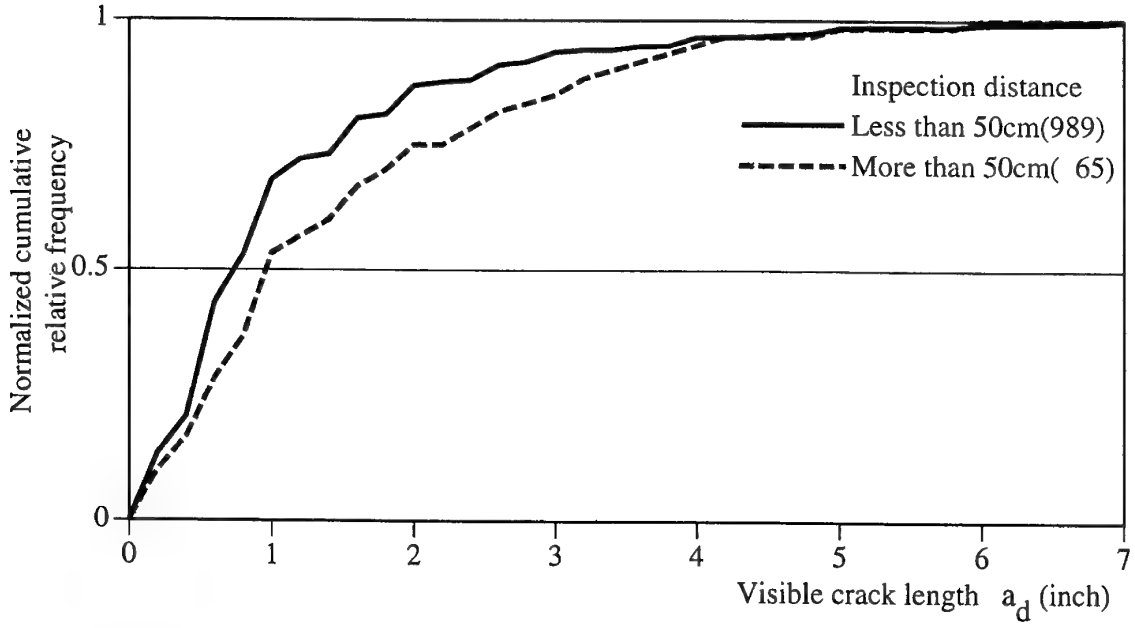


Figure 11 Influence of inspection distance on detected crack length

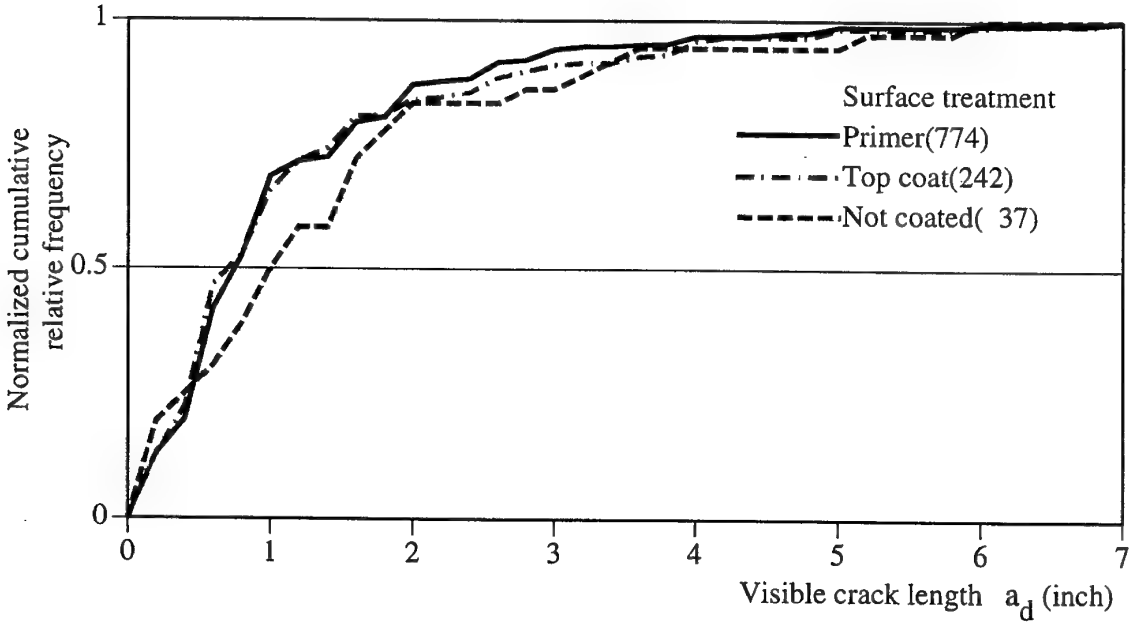


Figure 12 Influence of surface treatment on detected crack length

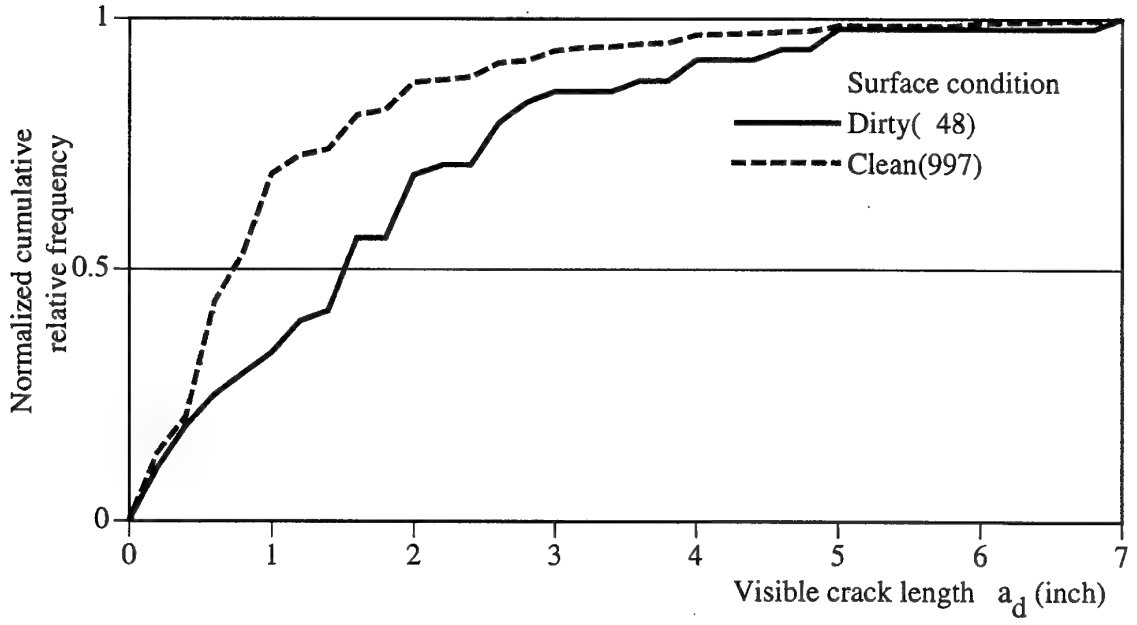


Figure 13 Influence of surface condition on detected crack length

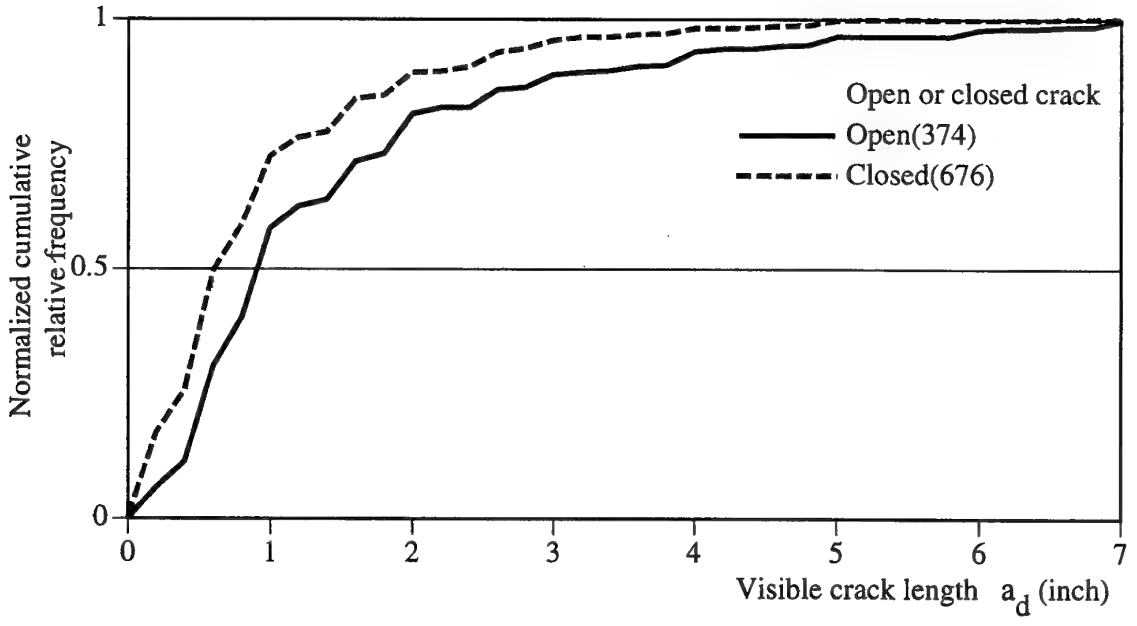


Figure 14 Influence of open or closed crack on detected crack length

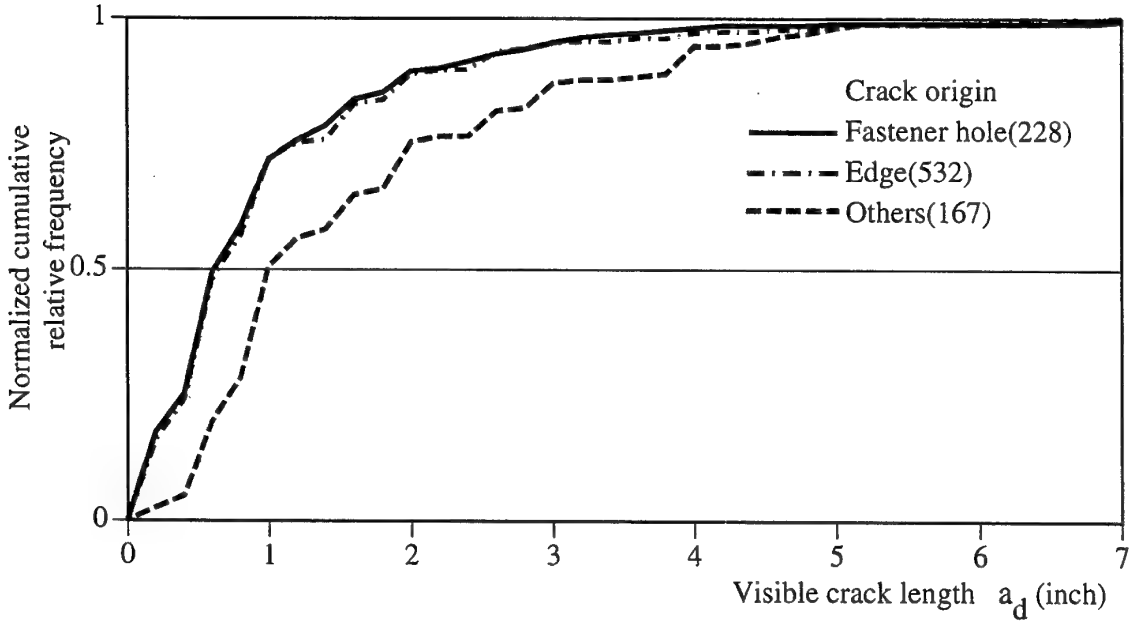


Figure 15 Influence of crack origin on detected crack length

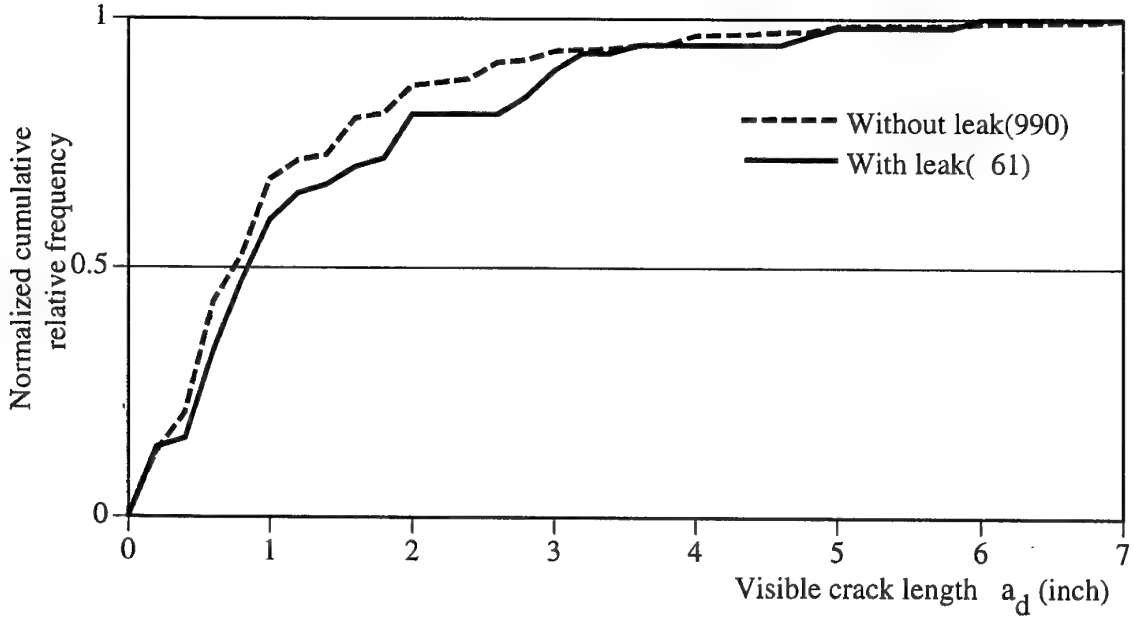


Figure 16 Influence of leak indication on detected crack length

Field Inspection Results and Damage Analysis of F-4F Horizontal Stabilizer Internal Structure

Jack T. Flood
Reinald E. Pfau
The Boeing Company
P.O. Box 516
St. Louis, MO 63166-0516
USA

Ernst Grauvogl
Friedrich Regler
Daimler-Benz Aerospace
Military Aircraft
85077 Manching
Germany

1. Summary

The McDonnell Douglas F-4F Phantom will remain in the German Luftwaffe inventory well beyond the year 2000. With the extensive usage in airforces all over the world, structural inspection programs based on fatigue tests and equally important, usage experiences shared with other countries provide a good knowledge of structurally critical areas of the airframe. However, depot inspections of the horizontal stabilizers discovered fatigue cracks in a rib that required fleet wide inspection through removal skin fasteners using boroscope and eddy current technique, performed by different "field-inspection teams" from the German Luftwaffe and industry.

Mathematical modeling of the local stress distribution with the damages zone together with periodic inspection provided the background for continuous A/C operation with damaged ribs and scheduled the sequence for replacements of cracked items.

A database for the reliability evaluation was gained by performing additional inspections on an original build-up structure with cracked ribs under in-field conditions.

2. Background

In August 1995 a crack was detected by the German Luftwaffe in a F-4 Phantom stabilator rib. Two months later additional cracks were detected by the industry in other aircrafts'. All cracks started at the same location. In order not to jeopardize flight safety, immediate action had to be taken. A risk analysis was carried out and

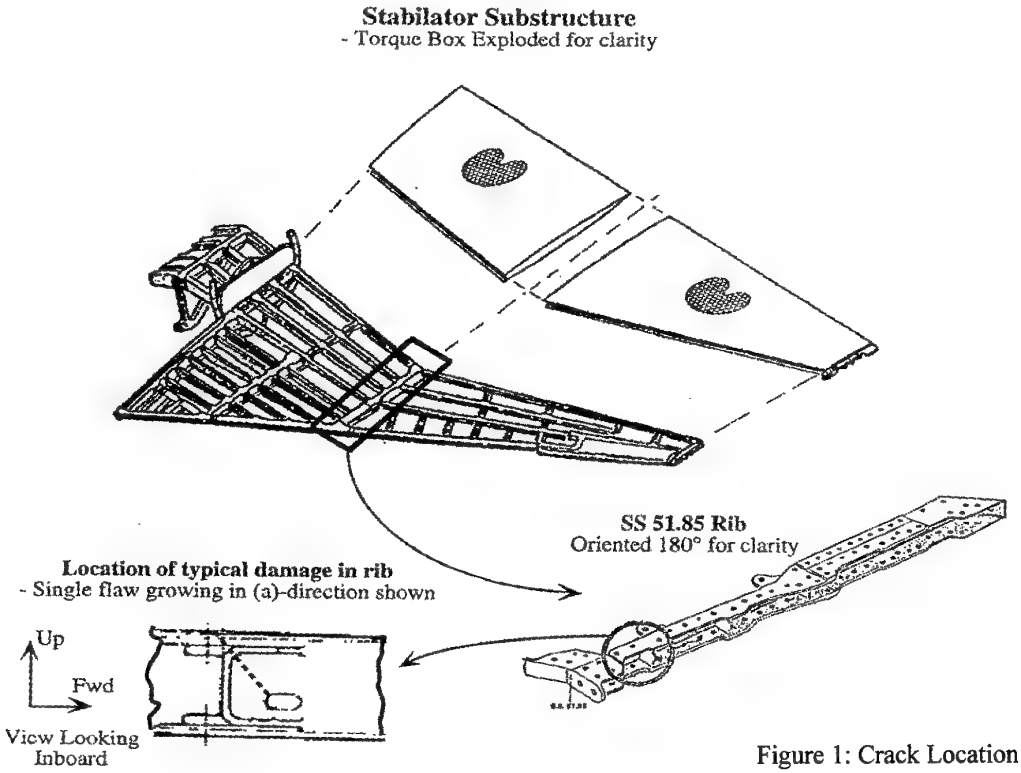


Figure 1: Crack Location

Paper presented at the RTO AVT Workshop on "Airframe Inspection Reliability under Field/Depot Conditions", held in Brussels, Belgium, 13-14 May 1998, and published in RTO MP-10.

the ribs were rated in four different categories:

- Cat. 1: no crack detected
- Cat. 2: cracks within established limits
- Cat. 3 repair necessary
- Cat. 4: rib has to be replaced

The challenge at that time was to provide enough engineering evidence to establish a sound decision. Therefore further in-depth analysis were needed to reduce the conservatism in the analysis approach and to lead the way for a final robust settlement. For a strong analysis evaluation it was crucial to obtain the in-field inspection data as soon and as error free as possible. The POD as well as the accuracy of the readings itself played a vital role in the analysis assessment.

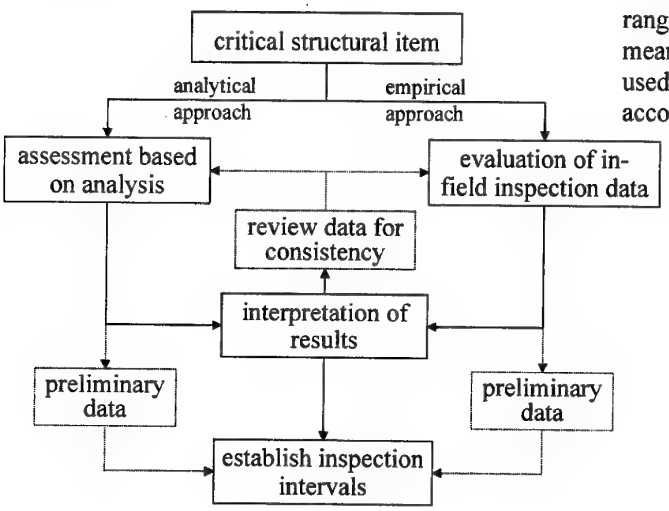


Figure 2: Principle Routes to Determine Inspection Interval

3. Analysis Approach

General

The analysis completed using the conventional maneuver load spectrum, developed for typical usage with the Modified Wheeler crack growth model, was not able to produce results that correlated with the fleet failures reported by the German Luftwaffe. In addition to the typical maneuver-loading environment, a review of flight test strain gauge data revealed that the stabilator structure experienced significant dynamic loading, during normal maneuvering, due to buffet effects. Buffet loads are considerably smaller than maneuver loads, but can occur with a frequency two three hundred times that of the maneuver loads. When buffet effects were included in the spectrum, fleet failures reported by the Luftwaffe could be predicted using the Modified Wheeler crack growth model.

Spectrum Development

The flight test aircraft was equipped with strain gage instrumentation. The measured buffet loads were presented in terms of root-mean-square (RMS) levels, power spectral density (PDS) functions, and statistical frequency parameters useful for fatigue analysis. Bending moment data from the flight test was measured at the times when the mean and RMS values were at their respective maximums in each of the eleven wind-up turn maneuvers performed with the aircraft in a clean configuration. The bending moment data values were normalized with the limit bending moment value and divided into ranges of mean values for both the maximum mean and the maximum peak. These values were used to modify the peak/valley data pairs to account for the dynamic loading influence.

In order to incorporate the peak/valley modifications, a Rayleigh distribution was used to define the probability for having a peak and a valley at certain loading levels. The cycle peaks due to buffet loading tend to follow a Rayleigh distribution. The overall amplitude of this distribution can be expressed as a function of RMS. The RMS values for the eleven wind-up turns maneuvers were separated into the ranges described above, and blocked to keep the number of spectrum load levels manageable. The blocking concept is illustrated in the figure below.

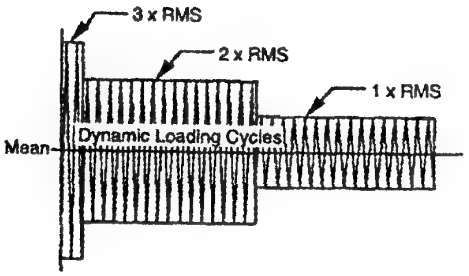
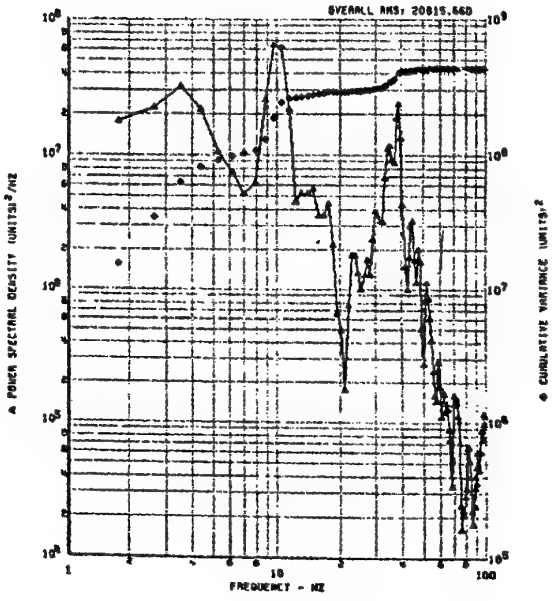


Figure 3: Blocked Rayleigh Distribution

The distribution average frequency is determined by calculating the centroid of the Power Spectral Density (PSD) versus Frequency curve, as shown on next page.

Figure 4: PSD Versus Frequency Curve



The centroid of the stabilator PSD curve was 22.4HZ. The average maneuver duration established during the F-4 ASIP buffet studies is five seconds. The total number of cycles for a given maneuver is determined by the product of the frequency and the duration, in this case 112 cycles. The dynamic cycles for a given maneuver can be established by imposing the blocked Rayleigh distribution on 112 cycles, multiplying each cycle to the static maneuver strain level. This process accommodates the magnitude of the dynamic effects on the peaks/valleys and the applicable number of cycles, such that the dynamic effects could be introduced into the existing maneuver spectrum. In order to incorporate the dynamic loading cycles into the spectrum, FORTRAN programs were written for the following three purposes:

a) to create a random spectrum from the block diagram
b) to calculate the effects of dynamic cycles using a statistical distribution
c) to incorporate the resulting effects on the spectrum peaks/valleys, and to modify the number of occurrences to accommodate these effects, as described above

Due to the fact that the flight test data was generated for a different stabilator location the programs were modified to allow for the effects of the buffet to be modified to calibrate the loading spectrum for the critical location. This was completed using the fatigue analysis to match the data gathered from the German F-4 fleet. Details of the fatigue analysis used to match the fleet data are presented later.

Finite Element Analysis

In order to idealize the localized rib structure and determine the applicable stress levels found in the area of interest, the existing non-slotted stabilator finite element model (FEM) was modified to more accurately depict the stress gradients present in the region. A view of the modified overall mesh geometry is shown below.

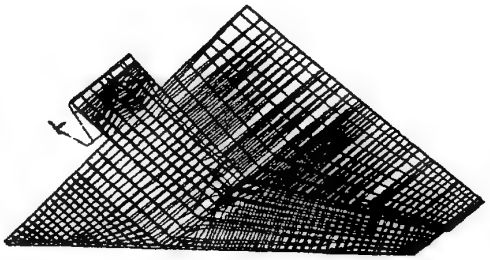


Figure 5: Finite Element Model

Extensive modification were also completed to the critical rib itself to more accurately depict the localized geometry and to refine the existing mesh in order to capture the detailed stress gradients required for analysis of the region. A portion of the refinements to the rib are shown, with a comparison to the original rib mesh, in the next figure.

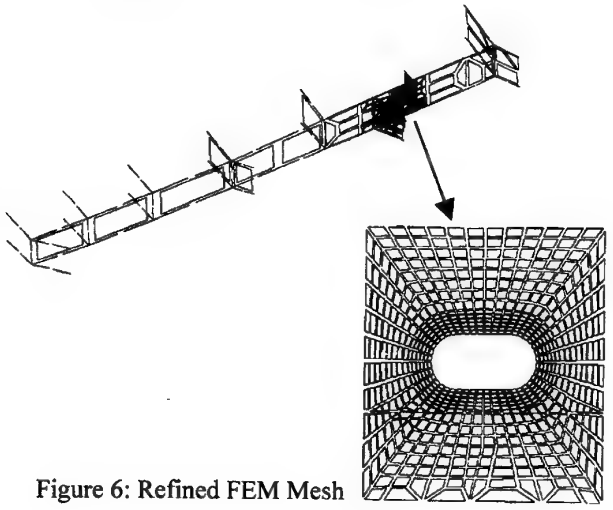


Figure 6: Refined FEM Mesh

The refined finite element model configurations were used to obtain the applicable stress gradients from the drain hole region in order to calculate stress levels and stress intensities for the crack growth analysis. Contour plots of the maximum principal stress levels were plotted, from the various model configurations in the drain hole region. An example of the stress gradients found in the region is displayed in the following figure, which contains the contour plot of the drain hole region.

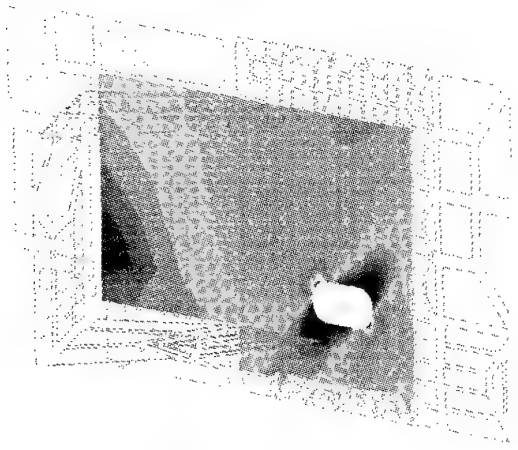


Figure 7: Contour Plot of Rib Web

Fatigue Analysis Overview

The F-4 fracture limit has been defined as the crack growth life for a flaw growing from a worst case initial flaw to failure. A study was performed to establish a conservative flaw size for single and double flaw scenarios. The study identified initial flaw sizes large enough to be statistically unlikely to occur in a manufactured hole, in order to establish initial flaw sizes from which conservative fracture limits would be calculated. It was determined that a 0.03" initial flaw for a single crack and a 0.01" initial flaw for a double crack would be conservative.

These parameters were implemented in the crack growth analysis of the stabilator rib. In order to identify the most conservative prediction for the different stabilator configurations, both initial flaw scenarios were analyzed.

Per the initial request by the German Luftwaffe, the crack growth analyses were to be predicted for the time to reach a crack length that leads to a rib replacement. To clarify this bound, discussions were completed by the German Luftwaffe and the industry and it was determined that a modification of this criteria would be required. In order to establish an acceptable bound, the analyses were completed to predict the crack growth life to reach the critical region in the upper stabilator skin, as determined by ultimate static strength checks. Negative strength margins were calculated in the upper skin, due to a crack having grown through the web and the upper rib flange from the drain hole. This bound was used as the fracture limit for all crack growth analysis completed for this report.

The crack length versus flight hours data from the German Luftwaffe was used to establish the proper input parameters for the crack growth runs which were used to calculate the predictions of the various baseline and repaired configurations. The match of this data was also used to calibrate the level of dynamic loading severity incorporated into the loading spectrum, as discussed earlier in the spectrum development section of this report.

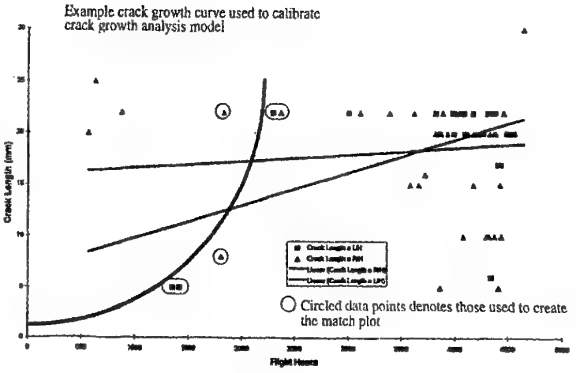


Figure 8: Crack Length Data Gathered In-Field

This data was incorporated to generate the input parameters for the crack growth propagating through the rib web from the drain hole. The proposed crack growth curve is included in the above figure. A conservative analysis would calculate a crack growth life in the vicinity of these data points and establish a set of parameters to be incorporated in the crack growth predictions for the various baseline and repaired configurations.

This match of the data was completed using the stress levels from the finite element model. A double flaw scenario was incorporated and the analysis was designed to predict the crack growth life in the web up to a length of 0.90", to match the fleet data. The figure below shows the results of the crack growth analysis for a single .03" through flaw, starting from the drain hole and growing through the web. The lower curve

shows the crack growth through the flange with an initial 0.07" flaw.

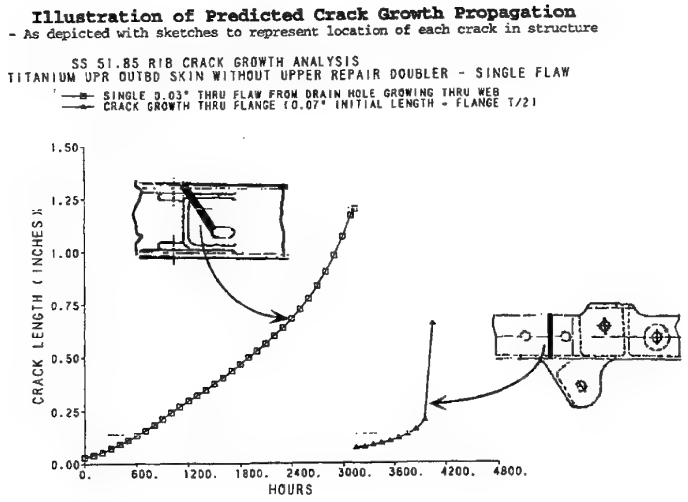


Figure 9: Crack Length vs. FH thru Web and Flange

With the validated crack growth curves established, the conservative inspection intervals were removed and new intervals were established by the results of the detailed analysis.

4. On Aircraft Inspection

In 1995 and 1996 on the GAF F-4F horizontal stabilators many cracks in aft end vertical legs were detected. Under field conditions these ribs were inspected with eddy curent (EC) high frequency inspection supported by visual inspection with a videoscope system. Access for the EC - probe and for the camera probe was given through bore holes.
Due to the occuring of many cracks in the F-4F horizontal stabilator rib 32-21122 the german F-4F fleet had to inspected urgently. This cracks were first found during the depot level maitenance by removing the stabilator skin starting from a drain hole and orientated in different directions. Alarmed by similar cracks in nearly every other stabilator in german maintenace facilities, the German Air Force (GAF) decided, to inspect every aircraft at short notice. Since the existing inspection technique was not able to test without removed skin, a new non-destructive inspection technique has to be developed quickly in order to detect and to quantify the cracks. Depending to the crack length and crack orientation the flight time and interval of inspection were determined. The inspection was finally performed with high frequency eddy current testing supported by a videoscope through removed fasteners. The eddy current probe was an especially developed pin probe (diameter 3mm) with special electric properties and an adapted shape.
The steering of the eddy current probe was monitored via a videoscope with 6mm outer diameter and 90° side view. It took some time for the inspectors to get used to this „remote" inspection and also to estimate the crack length. To ease this, a sketch with orientations and maximum crack length were given. (Figure 10).

For the inspection on aircraft three fasters were removed: one for the eddy current probe, one for the monitoring videoscope and one for an additional videoscope inspection. The in-field inspection were performed by an inspection team of Daimler-Benz Aerospace and German Air Force. A standarized inspection report was to be reported and additional a video printout of detected cracks was made.
The inspection has to be performed in hangars (Figure 11) or on airfield at nearly every weather. The area of inpection was mostly dirty and loose foreign object fasteners were fixed in the drain hole, where removal of this fasteners was not always possible. This results in a limited inspection access and a limited inspection result. Due to the short notice of the inspection the NDI-teams were on under high pressure.

5. Post Analysis of the Inspection Results

In the mean time most of the cracked ribs are replaced and a reliabiltity evaluation on the replaced ribs was done. Several inspectors were involved in this progress. The cracked ribs were conditioned in an orginal build - up and the inspectors had to fulfil their job under PDM (periodic depot maintenance) conditions. Also inspection results from the earlier inspections were considered in the reliability evaluation.

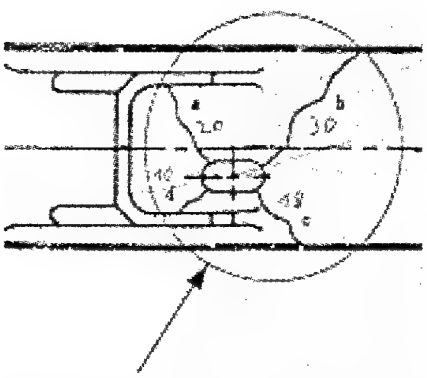


Figure 10: Area of Inspection and Predefined Crack Length

Depending on the results of the initial inspection some stabilators were able to fly without any limitations, some were re-inspected within an interval of 50 flight hours and some were brought direct to repair facilites, where the ribs with cracks were replaced. In this way many ribs with cracks were available and give the chance to check the inspection reliability of this difficult inspection problem. For this reason seven inspectors with more or less experience about this particular inspection have to fulfill their job again on a dummy set-up in order to give nearly the original inspection contitions (Figure 12). All the inspectors are obtaining a qualification according to prEN4179, respectively DIN 65450, which is equivalent to MIL Std. 410, with different times of experience (Figure 13).
Like under aircraft inspection conditions the inspectors had access to an original rib with maximum crack length in mm to estimate the crack length and the orientation. Due to this many results with maximum crack length are showing very similar results.
By the 7 inspectors18 ribs with 2 areas of inspection were tested and evaluated giving a total number of 252 measurement points. In the reliability study five categories of findings were represented because of its influence on further aircraft treatment (i.e. flight without limitations, inspection interval, removal of stabilator):

- False Call
- Crack not found
- Crack found, but measurement too low (>+-10%)
- Crack found, but measurement too high (>+-10%)
- Correct findings (including also correct inspection of none cracked area)

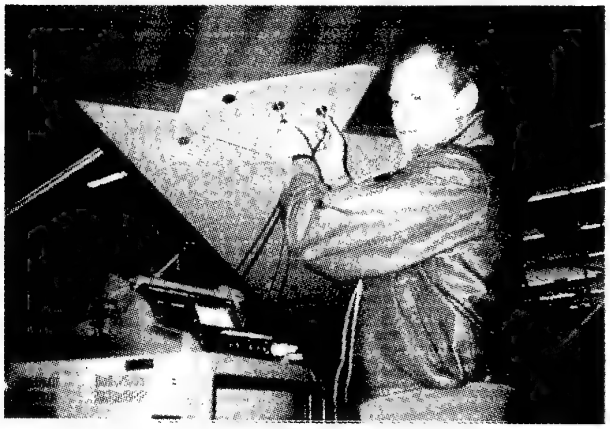


Figure 11: Inspection of F-4F Internal Structure on Aircraft

The results of this post analysis is shown in Figure 14. It shows that some cracks were not found which are mostly small ones. A slightly increasing number of indications were false calls which would mostly result in unneccesary reinspection of the aircraft. The number of measurements where cracks are measured too small are significant lower than that with cracks measured too high. This is a positive tendency for aircraft safety but can result in higher inspection costs because of more reinspected aircraft. Nevertheless the really obvious result of this analysis is that most of the ribs were correct inspected and the very most of the cracks were found with the more or less correct crack length.
Regarding only the correct findings of the cracks leeds to a Probability of Detection (95% Confidence Level) shown in Figure 15. This chart shows for example, that cracks with 6mm length were found at a 80% probability.
The inspection results obtianed during on aircraft inspection where available of most of the inspected ribs. This result were also compared with the „real defects“ tested prior to the simulated inspection under ideal conditions also with eddy current. This comparison shows a nearly similar behaviour like the simulated inspection beside a significant number of missed cracks (Figure 16). This might be mainly due to the fact, that there is an undefined period of flights between

the on aircraft inspection and the verification. This is also true for cracks with values too low and the correct findings. False calls and cracks with values to high are definetly wrong measurements. Due to the undefined times of inspections a POD analysis was not reasonable. This fact leeds to call for a similar reliability study but with removal of the inspected part direct or shortly after the inspection on aircraft.

Figure 12: Dummy Set-up for Simulated Inspection

Inspector	ET	UT	PT	MT	RT	LEAR	Time of experience Eddy Cuurent Testing (years)
A	1	1					1
B	2	2	2	2	2	X	8
C	2	2	2	2	2	X	12
D	2	2	2	2	2	X	8
E	1	1	1	1	1		4
F	2	2	2	2	2		4
G	1	1	1	2	3		2

Figure 13: NDI-Personnel Qualification and Approval for Eddy Current Testing (ET)

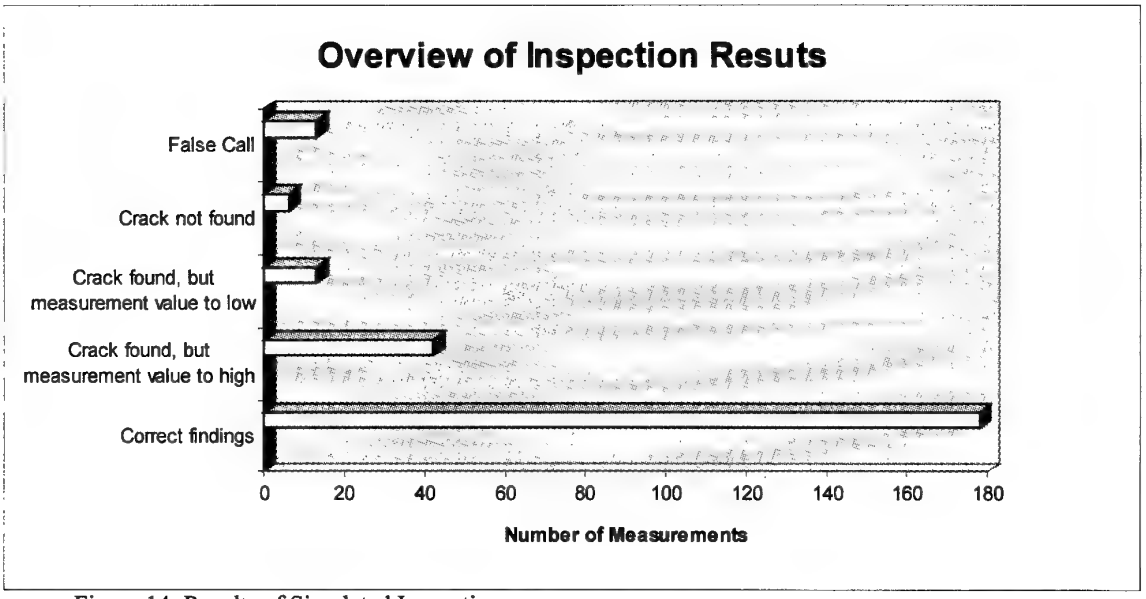


Figure 14: Results of Simulated Inspection

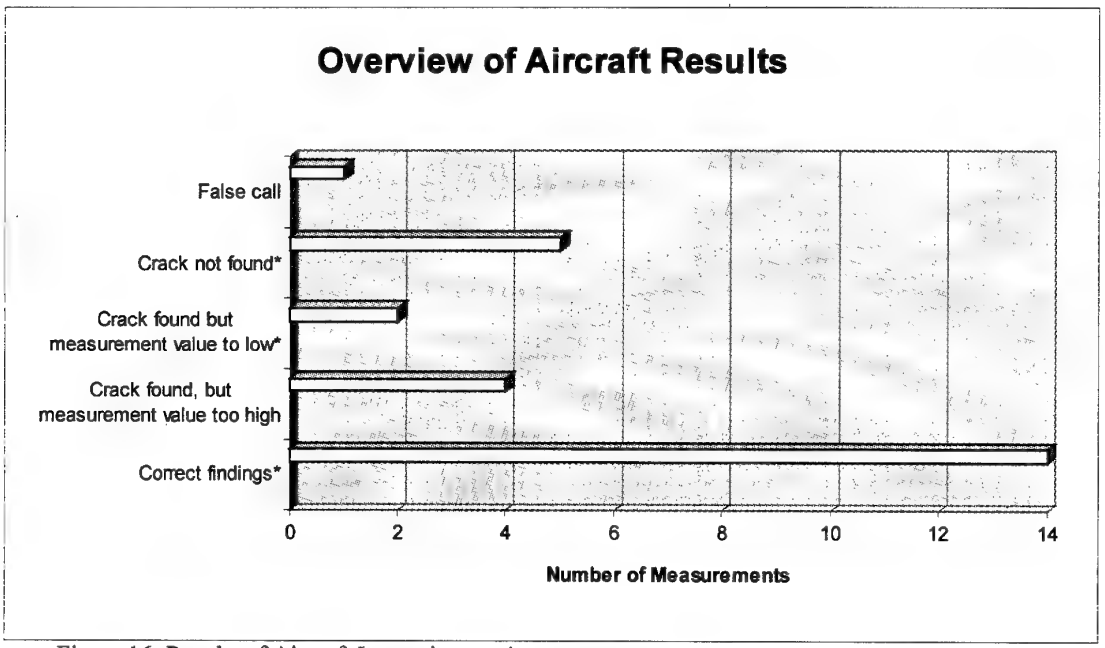


Figure 16: Results of Aircraft Inspection * **Undefined flight hours between measurements can increase cracks**

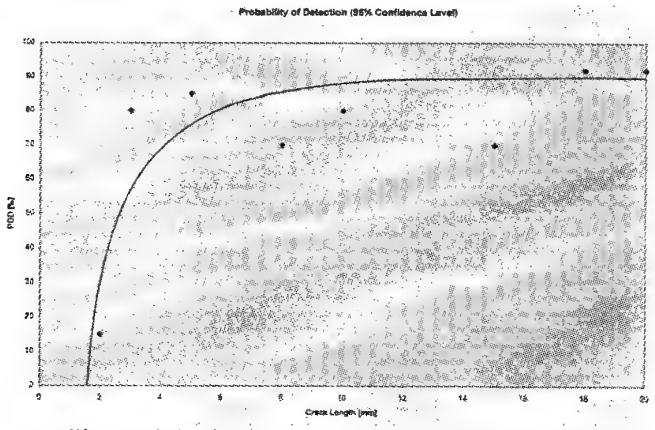


Figure 15: Probability of Detection (POD) of the Simulated Inspection

DOUBLE PASS RETROFLECTION VERSUS VISUAL INSPECTION OF IMPACT DAMAGE IN COMPOSITES

D. S. Forsyth
R. W. Gould
J. P. Komorowski
Institute for Aerospace Research
National Research Council
Building M14, 1191 Montreal Road, Ottawa ON Canada K1A 0R6
email jerzy.komorowski@nrc.ca

SUMMARY

Double Pass Retroreflection is the basis of the D Sight™ Aircraft Inspection System, (DAIS). The DAIS 500 equipment is built specifically for the rapid inspection of large composite surfaces for impact damage. In one sensor placement 0.27 m^2 can be assessed for the presence of impact damage.

Impact indentations of 0.025 mm depth are observed with high rates of detection in D Sight™ inspections of certain composite structures. Current structures are designed with generally accepted BVID (barely visible impact damage) limits of between 1 to 2.5 mm deep dents. Benefits which could be derived from the implementation of DAIS to current and future composite structures inspections are briefly discussed.

For damage tolerance design purposes, probability of detection (POD) data for DAIS equipment is required. This report presents inspection results from a set of inspectors with different levels of training, with two sets of specimens, one derived from a vertical stabiliser surface area on the CF-18, and one derived from hat stiffened composite panels built at the Institute for Aerospace Research (IAR).

It is shown that inspectors, almost independent of the training time, can reliably identify impact indents an order of magnitude better than currently accepted BVID limits, with almost zero false calls. This sensitivity and reliability, when combined with low cost and speed of DAIS inspections, demonstrate that the DAIS 500 is an excellent tool for rapid, wide area inspection of composite structures for BVID.

LIST OF SYMBOLS

BVID – barely visible impact damage
DAIS – D Sight™ Aircraft Inspection System
IAR – Institute for Aerospace Research
NDI – non-destructive inspection
NRC – National Research Council (Canada)
POD – probability of detection

1. IMPACT DAMAGE IN COMPOSITES

The use of graphite reinforced resins is increasing in airframes of military and civil aircraft. These materials offer high specific strength and stiffness properties and very good fatigue resistance. Unfortunately, the materials are sensitive to low energy impact damage from such common occurrences as hailstones, stones thrown off the runway, or tools dropped by maintenance personnel. These impacts may result in significant levels of internal damage while surface damage may be barely or non-visible.

One study of operational experience with composite structures indicated that 81% of all damage found was due to impact while lightning strikes (10%), overheating (7%), and delamination (2%) constituted the remainder of damage types (Ref. 1).

Regular in-service inspections of aircraft with scanning devices are not practical due to the cost and time required. Currently, operators rely on visual inspection for impact damage. Different organisations have assumed different thresholds of detectability for impact damage, with little published data supporting these thresholds.

Damage tolerance requirements state that composite aircraft structures should be capable of carrying the design ultimate loads after sustaining impact damage below the detectable size limit. The United States Air Force (USAF) Damage Tolerance Design Guide for composites defines this limit for visible impact damage as a 2.5 mm (0.1 inch) deep indentation. Other organisations have established lower thresholds (Ref. 2): The United States Navy uses 1.25 mm or 0.05 inch, the US Federal Aviation Administration uses the limit of detectable damage, while Aerospatiale of France has used 0.3 mm (0.012 inch) as the visibility threshold (close visual inspection with 50% probability of detection for the ATR 72 composite wing box). The structural repair manual for the CF-18 (Ref. 3) specifies a limit of 0.125 mm (0.005 inch) deep damage on the vertical stabiliser before repair action must be taken, with the first line inspection being a visual inspection. These attempts to lower the threshold are driven by the desire to design lighter structures with higher allowable strain levels. There is no evidence to substantiate that typical visual inspection can reliably detect impact damage of as small as 0.125 mm.

Research at IAR (Ref. 4) and at Aerospatiale (Ref. 5) has shown that significant relaxation can occur at impact damage sites, resulting in reductions (up to 45%) in impact dent depths. This can be caused by viscoelastic effects, cyclic loading, moisture, and temperature effects. This implies that if visual inspections are to be used, then higher impact energies required to produce visible damage after relaxation (i.e. a residual indent depth equal to the required value) will be needed for certification. As a consequence the allowable design strain levels will have to be lowered even further.

A cost effective method for rapid, regular, inspection of composite structures with a capability better than close visual inspection would not only reverse this requirement for higher impact damage and lower strain allowables, but offers the potential to lower the impact requirement and increase design allowables (Ref. 6). This would result in lighter composite structures and also enhance safety of operation of current

Paper presented at the RTO AVT Workshop on "Airframe Inspection Reliability under Field/Depot Conditions", held in Brussels, Belgium, 13-14 May 1998, and published in RTO MP-10.

designs where relaxation was not accounted for during certification.

2. D SIGHT™ AIRCRAFT INSPECTION SYSTEM

In 1988, Komorowski and Gould of NRC suggested the development of optical impact detection systems based on double pass retroreflection, also know as D Sight™ (Ref. 7). The Canadian Department of National Defence and US Air Force joined Diffracto Ltd. and NRC in sponsoring development of a commercial D Sight™ Aircraft Inspection System (DAIS 500) for impact damage detection (Ref. 8). Several of these systems have been delivered to the USAF and the Canadian Air Force.

The DAIS systems consist of an inspection head containing the optics, CCD camera and light source; a personal computer running DAIS software with inspection planning, acquisition, analysis and repair modules; and remote pendant with touch sensitive screen for controlling the acquisition process. DAIS requires two operators. The first operator is responsible for placing the inspection head on the surface of the aircraft. The second, the pendant controller, uses preplanned placements shown in the pendant to direct the first operator. Once the required position is achieved, the pendant controller uses the touch screen to save the D Sight™ image.

The acquired images are shown on the pendant touch screen. It is possible for the pendant controller to make an immediate assessment, but the recommended procedure is to postpone the image analysis and complete the acquisition process for the whole inspected surface or complete aircraft. This reduces the time during which the aircraft has to be available to the inspectors. Also, image analysis is better carried out at a workstation using a CRT monitor (the DAIS PC and pendant are equipped with LCD screens). Results of the analysis are reported on wire-frame diagrams of the aircraft type and can easily be referenced back to locate the detected damage (Ref. 9).

The only factor that may be influenced by the operators during the acquisition part of the inspection process is the surface reflectivity. The D Sight™ process requires that the light be reflected twice from the inspected surface. Most military aircraft are coated with matte, non-reflective paint. To allow DAIS inspection these surfaces are temporarily wetted with a highlighting fluid. The operators must ensure that the highlighting is sufficient and uniform. The DAIS informs the operators when the average reflectivity is below a required minimum and automatically adjusts the light intensity to produce consistent images. The inspector performing the image analysis will easily spot non-uniform highlighting and can request that the affected images be reacquired. Solid film highlighting has recently been developed (Ref. 10). This approach removes reflectivity as a variable in the acquisition process. However, solid film highlighting has not yet been incorporated into commercial DAIS systems.

3. EXPERIMENT

3.1 Non-destructive Inspection Reliability Experiments

The design of reliability experiments for non-destructive inspection (NDI) has been well documented (Ref. 11, 12). At the design stage, care must be taken to ensure that the experiments to be performed can resolve the desired variables. The expense of developing specimens and conducting inspections for POD trials can be prohibitive, and the number of experiments that must be performed increases geometrically with each variable being investigated.

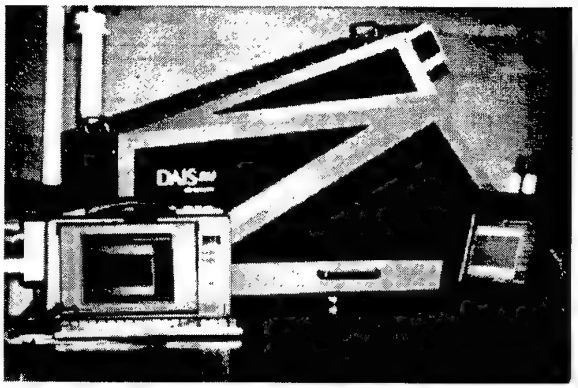


Figure 1. The DAIS 500, with the inspection head in the background and pendant on the right.

A set of experiments was designed to determine the POD which can be attained by inspectors with different levels of experience under a constant set of external physical conditions. The operation of the DAIS 500 unit was not a variable in this study, as each inspector was presented with the same set of images from DAIS 500 inspections of physical specimens.

A significant amount of practical information can be determined from this experiment design. The effect of training and experience on interpretation of DAIS 500 images can be estimated. The sensitivity of the DAIS 500 inspections can also be estimated for impact damage flaws on the types of specimens inspected herein.

3.2 Experimental Data Sets

One of the important factors in NDI reliability experiments is cost. In most cases, it is very expensive to generate the numbers of test specimens which are required to give statistically valid results. Because of the simplicity of execution of the DAIS inspection, variability in the data acquisition process is not a significant factor in the reliability of the inspections. Considering the nature of the DAIS inspection technique, it was possible to artificially generate the data sets in this experiment from a small number of inspections of actual specimens.

Based on the assumption of a repeatable data acquisition process, the data sets used in this work were generated by seeding a small number of images with different size flaws in different locations on the DAIS image. One data set was based on DAIS inspections of hat stiffened composite panels built at the Institute for Aerospace Research. The other data set was from inspections of a vertical stabiliser surface area on the CF-18.

As part of the development of DAIS as a large area composite inspection system, four hat-stiffened panels had been impacted at various energies and locations with a 12.5 mm (0.5 inch) spherical indentor. Some of these panels had over 50 impact sites. Each panel was inspected with the DAIS 500 such that the impact site was collected in the top, middle and bottom third of the view. This is important because the DAIS 500 image is skewed in the vertical direction, and the sensitivity and skewness depend on the vertical position in the image. In creating test images, flaws were positioned in the same vertical third of the new image as they originally appeared in the master image. Thus the skewness and sensitivity changes were minimised. Six stiffened panels, of size 75 x 90 cm (30 x 36 inches) had not been subjected to the impact tests and their surfaces were imaged at 4 different locations each. Forty images

were also collected, at various locations, from an undamaged CF-18 vertical stabiliser.

From the 178 impact damage sites available, sets of candidate sites were cut from the original view and placed in the corresponding third of an undamaged 'background' image. The cut and paste procedure was carried out manually with the use of graphics software. For images with a damage site, the x and y co-ordinates of the placement were recorded and used to adjudicate the responses of the inspectors.

Two sets of 100 images were created: one based on the stiffened panels and one on the CF-18 vertical stabiliser. In each case the undamaged surface 'background' images were used for both the 50 damaged and 50 undamaged images in each set.

3.2.1 Hat-Stiffened Panels

The hat-stiffened panels were made in two different lay-ups using two material systems. The first generation prepreg material was unidirectional carbon fibres pre-impregnated with epoxy resin, Hercules 3501-6. The carbon fibres were Hercules Magnamite continuous type AS4. This material has been widely used in the aerospace industry for over 15 years and is the material used on the Canadian CF-18 fighter aircraft.

The other composite material system selected for this study was unidirectional carbon fibres pre-impregnated with bismaleimide resin, Cytec's Rigidite 5250-4. The carbon fibres were Hercules Magnamite continuous type IM7.

Two hat-stiffened panels 75 x 90 cm with different lay up configurations were designed (see Table 1). Configuration 1 was designed as a lightly loaded fairing type structure while configuration 2 was a heavily loaded wing skin type structure. Figure 2 shows a photograph of a typical panel of configuration 1.

Table 1. Lay-ups for the two hat-stiffened panel configurations.

			Laminate Thickness (mm)	
			IM7/ 5250-4	AS4/ 3501-6
configuration 1				
	plies			
skin	12	$(45/0/45/90)_s$	1.53	1.58
cap	26	$(45/0_4/90/0_3/45/0)_s$	3.30	3.43
web	12	$(45/90/0/45)_s$	1.53	1.58
flange	5	(45/90/45)	0.64	0.66
configuration 2				
	plies			
skin	48	$(45/0/45/90)_{4s}$	6.10	6.34
cap	52	$(45/0_4/90/0_3/45/0)_{2s}$	6.60	6.87
web	24	$(45/90/0/45)_{2s}$	3.05	3.17
flange	10	$(45/90/45)_2$	1.27	1.32

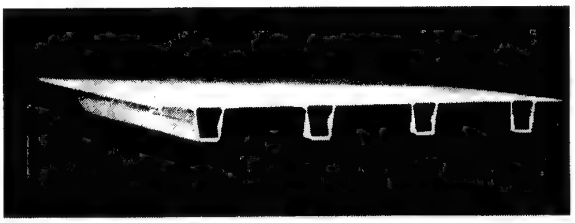


Figure 2. Configuration 1 hat-stiffened panel.

3.2.2 The CF-18 Vertical Stabiliser

As mentioned above, the material system used for the skin of the CF-18 vertical stabiliser is composed of a prepreg material of unidirectional carbon fibres pre-impregnated with epoxy resin, Hercules 3501-6, and carbon fibres of Hercules Magnamite continuous type AS4.

The skin of the vertical stabiliser is made up of a number of panels of different thicknesses and number of plies, in a quasi-isotropic lay-up. Thicknesses range from 5.6 mm to 1.7 mm. The repair manual states that dents greater than 0.005 inch (0.125 mm) in depth be repaired.

3.3 Inspection Procedures

A number of inspectors with different levels of NDI experience and training evaluated both sets of DAIS images. Only one inspector had DAIS experience. The inspectors were provided with a reference image for each data set, with labelled flaw locations. Figure 3 shows the reference image for the set of hat stiffened composite panels.

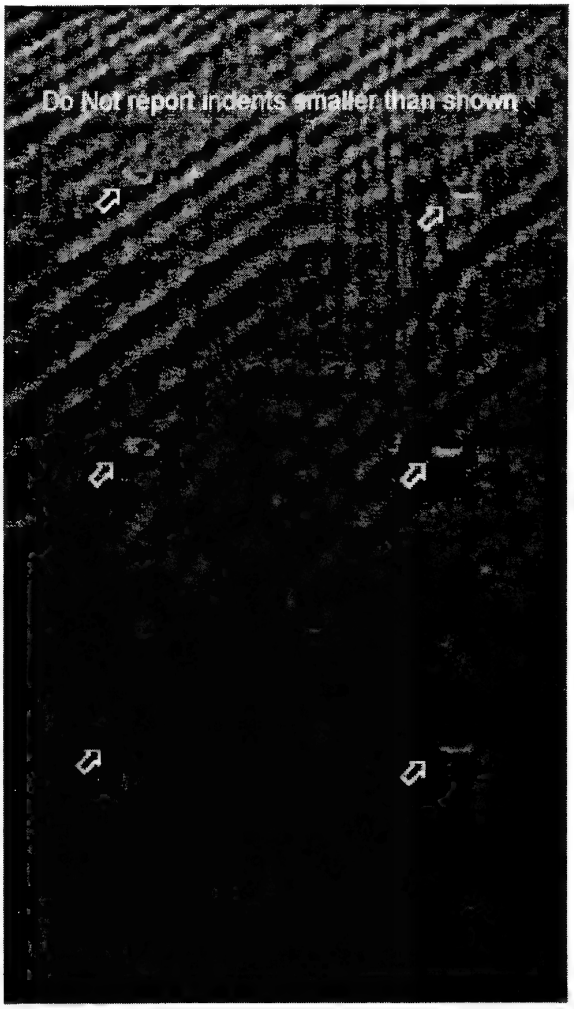


Figure 3. The DAIS 500 reference image for the hat-stiffened composite panels, showing minimum flaw sizes for inspectors.

The reference image showed the minimum flaw sizes the inspectors should note. Again because of the skewness and variable sensitivity of the DAIS image, the reference images

showed these minimum flaw sizes for the top, middle, and bottom thirds of the DAIS image.

The interpretation procedure was performed with the inspector sitting in front of a computer, looking at both the reference image and the test image. Minimum computer requirements were for a 15 inch diagonal monitor size and a 1024 by 768 pixel resolution in 256 colours. The inspector evaluated each image in order. If the inspector found a flaw site, the location in pixels was recorded. No estimation of flaw size was recorded. A typical example of a DAIS 500 inspection from the hat-stiffened panel data set is shown in Figure 4. An example of a DAIS 500 inspection of the CF-18 vertical stabiliser is shown in Figure 5.

Inspection results are recorded as "hits", "misses", or "false calls". A hit occurred when the inspector correctly located a flaw on an image. A miss occurred when an inspector did not find a flaw on an image. A false call occurred when an inspector incorrectly noted a location as being flawed. Note that a miss and false call can occur on the same image by these definitions.

Additional information recorded for each inspection was the inspector, the inspector's age, whether the inspector used corrective lenses, whether the inspector had any NDI experience, and total time to complete each data set.

Figure 4. An example of a DAIS 500 inspection of a hat-stiffened composite panel, with a 0.006" or 0.15 mm deep impact.

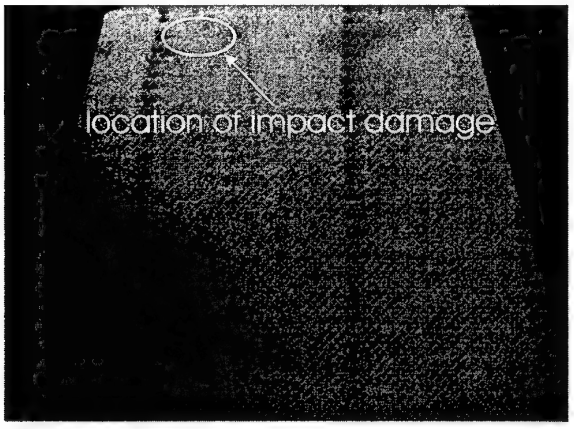


Figure 5. An example of a DAIS 500 inspection of a CF-18 vertical stabiliser, with a 0.001" or 0.025 mm deep impact.

4. RESULTS

Histogram-type presentations of the detection rate for the different impact depths are presented below. Figure 6 shows the detection rate for impact depths from 0.006" or 0.15 mm to 0.010" or 0.25 mm, in the hat-stiffened panels. The data is shown for all inspectors, and for inspectors with and without previous NDI experience. The 95% confidence bounds are shown on the data for all inspectors. Figure 7 shows the detection rates for all inspectors, with the data broken down to show differences due to location of the flaws with respect to the DAIS image. Detection rates averaged for all inspectors ranged from 0.85 with a 95% confidence range of +/- 0.02, to 1.00 with a 95% confidence range of +/- 0.00.

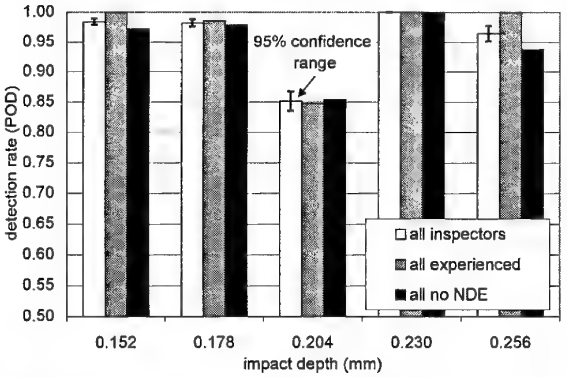


Figure 6. Detection rates for the DAIS 500 inspection of impact damage on hat-stiffened composite panels.

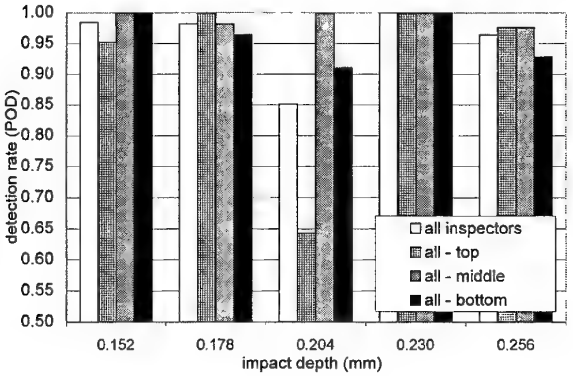


Figure 7. Detection rates for the DAIS 500 inspection of impact damage on hat-stiffened composite panels, organised by location of flaw with respect to the DAIS 500 image.

The surface of the CF-18 vertical stabiliser is more uniform than that of the stiffened panels, except for fasteners. A set of smaller impact damage sites were used for this inspection. Figure 8 shows the detection rate for impact depths from 0.001" or 0.025 mm to 0.005" or 0.13 mm, in the CF-18 specimens. The data is shown for all inspectors, and for inspectors with and without previous NDI experience. Figure 9 shows the detection rates for all inspectors, with the data broken down to show differences due to location of the flaws with respect to the DAIS image. Detection rates averaged for all inspectors ranged from 0.86 with a 95% confidence range of +/- 0.02, to 0.98 with a 95% confidence range of +/- 0.01.

The inspection results were also broken down by location of flaw with respect to the DAIS image. These results are shown in Figure 7 and Figure 9.

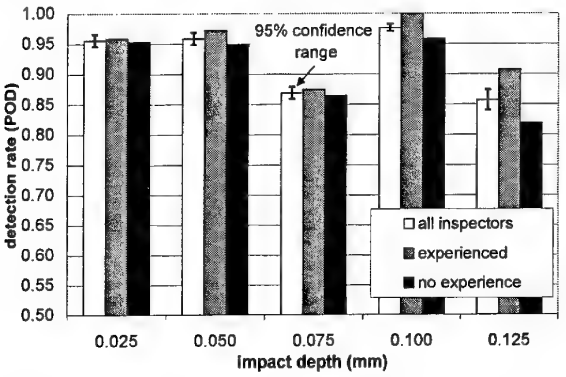


Figure 8. Detection rates for the DAIS 500 inspection of impact damage on a CF-18 vertical stabiliser.

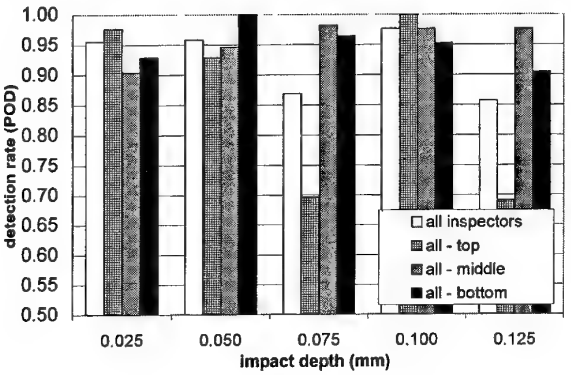


Figure 9. Detection rates for the DAIS 500 inspection of impact damage on a CF-18 vertical stabiliser, organised by location of flaw with respect to the DAIS 500 image.

Table 2 shows false call rates for each inspector, broken down to show differences between inspectors with and without NDI experience. Table 2 is for the inspections of the hat-stiffened composite panels.

Table 2. False call rates for inspections of the hat-stiffened composite panels.

	no NDI experience	with experience
false call rate (%)	0	2
	8	4
	4	8
	5	14
	2	3
	16	2
	5	N/A
	10	N/A
average	6.3%	5.5%
standard deviation	5.0%	4.7%

Table 3 shows false call rates for each inspector, broken down to show differences between inspectors with and without NDI experience. Table 3 is for the inspections of the CF-18 vertical stabiliser.

Table 3. False call rates for inspections of the CF-18 vertical stabiliser.

	no NDI experience	with experience
false call rate (%)	2	1
	7	0
	10	0
	0	13
	5	0
	6	1
	6	N/A
	1	N/A
average	4.6%	2.5%
standard deviation	3.4%	5.2%

5. DISCUSSION

The results of this study were analysed for the purposes of generating POD curves as a function of impact depth. However, in both the stiffened panel and the CF-18 vertical stabiliser, the data were not sufficient to generate POD curves. This was because over the range of impact sizes examined, there was little change in POD, and it is not practical to generate and measure impacts that are not detected by DAIS. Because the DAIS inspection yields varying sensitivity across the image, inspection results were also analysed for variations due to the flaw location within the DAIS image.

This data establishes that experienced NDI inspectors, some of whom have no training with the DAIS system, can achieve very high rates of detection on impact damage sites of 0.025 mm, depending on the roughness of the undamaged background. This is one to two orders of magnitude better than the BVID design limits quoted previously, for which the reliability of visual inspections are not substantiated by experimental data.

In both specimen types, the POD is not well correlated with impact size, over the range of impacts tested. There are a couple of factors which may contribute to this. First, the very high detection rates of very small flaws on the smooth CF-18 vertical stabiliser surface demonstrates that the sensitivity is very good. The background noise and resulting signal to noise ratio is probably the limiting factor. Fastener rows and other local surface features may reduce the POD for impacts in their immediate area. Secondly, the detection of impact sites by DAIS is due to the slope around the site, not the actual depth. The hemispherical indentor used in this test produces minimal slope changes for a particular impact depth, and the slope changes are not necessarily correlated to the impact depth.

Because the D Sight™ images are skewed, and of varying sensitivity across the image, results were also broken down by flaw location (see Figure 7 and Figure 9). While there is significant variation in detectability for some flaw sizes, there is no systematic variation due to location of flaw with respect to the DAIS image.

Another important consideration in the performance of an NDI system is false calls. While the false call rate does not affect POD, it is an economic issue, as false calls cause unnecessary costs of downtime, repair, or replacement.

There is no difference between the false call rates of experienced and inexperienced inspectors for the hat-stiffened panels. These panels had a rougher surface, which is equivalent to a lower signal to noise ratio for this inspection (see Figure 4). On the inspections of the smoother CF-18 vertical stabilisers (see Figure 5), the false call rates of the experienced inspectors

decreased from 5.5% to 2.5%, while the inexperienced inspectors showed only a slight decrease from 6.3% to 4.6%.

6. CONCLUSIONS

An unmodified commercially available DAIS 500 inspection system was used to inspect two composite structures for impact damage. Sets of inspection images were created by cutting and pasting flaw signatures from actual impacts on a number of images without flaws. Fourteen inspectors with different levels of NDI experience evaluated the resulting inspection images.

On hat-stiffened graphite-epoxy panels built at IAR, inspectors as a group achieved detection rates of between 85% and 100% on seeded impact damage flaws of depths between 0.152 mm (0.006") and 0.256 mm (0.012").

On the CF-18 vertical stabiliser, experienced inspectors achieved rates of detection between 88% and 100% on seeded flaws ranging in size from 0.025 mm (0.001") deep to 0.125 mm (0.005") deep.

Inspectors with NDI experience, though without DAIS training or experience except for one, outperformed inspectors with no NDI experience. This indicates that further DAIS specific training may improve results. IAR has also developed a solid film highlighter technique which provides constant illumination, and de-skewing algorithms to correct the DAIS image. These developments are expected to further improve the sensitivity of DAIS image interpretation.

Some previously published data exists on the POD of visual and other inspections for impact damage in composites, but care must be taken to make comparisons. In different material systems, and different lay-ups in any one material system, the energy to create a constant impact depth is different. If the same size and shape indentor is used, the relationship of impact depth to impact diameter is constant, even if different impact energies were used. For a damage-tolerance analysis, the measure of interest is impact energy. However, the detectability by visual or enhanced visual inspection of a flaw will be due to its size and shape with respect to the background surface roughness.

The CF-18 structural repair manual states that damage in the vertical stabiliser skins exceeding a depth of 0.005 inch (0.125 mm) must be repaired. The first line inspection is visual. While this level of damage is detectable using the DAIS 500 system, there is no evidence that this can be done visually in depot conditions. The use of this figure in the repair manual is thus of questionable value.

The USAF Damage Tolerance Design Guide for composites requires that a structure must sustain an impact flaw depth of 0.100 inch (2.5 mm), and the USN requirement is 0.050 inch (1.25 mm) (Ref. 2). These numbers are based on what is believed achievable by visual inspection (although this has not been demonstrated in published literature), and are much higher than what has been demonstrated for the DAIS system. If the DAIS inspection procedure was used as an alternate means of compliance, the USAF and USN requirements could be changed for these selected areas, to reflect the detectability of impact damage on individual material systems.

7. ACKNOWLEDGEMENTS

This study was carried out with the support of NRC and CRAD, Department of National Defence. The authors also would like to thank all the inspectors who took part in this study.

8. REFERENCES

1. Schur, F., "Inspection of Carbon Fibre Repairs", Air Transport Association Non Destructive Testing Forum, Long Beach, California, 1991.
2. Kan, H. P., Whitehead, R. S., and Kautz, E.,"Damage Tolerance Certification Methodology for Composite Structures", NASA CP 3087.
3. C-12-188-SRM/MM-005 Structure Repair: Aft Fuselage, CF-18 Hornet Library.
4. Komorowski, J. K., Gould, R. W., Marincak, A., "Study of the effect of time and load on impact damage visibility", Proceedings of the Second Canadian International Composites Conference and Exhibition (CANCOM 93), W. Wallace, R. Gauvin and S. V. Hoa Editors, Ottawa, Ont., September 1993, pp 441-446.
5. Thomas, M., "Study of the evolution of the dent depth due to an impact on carbon/epoxy laminates. Consequences on impact damage visibility and on in service inspection requirements for civil aircraft composite structures" presented at MIL-HDBK 17 meeting March 1994, Monterey, California.
6. Komorowski, J. K., Gould, R. W., and Simpson, D. L., "Synergy Between Advanced Composites and New NDI Methods", Advanced Performance Materials, **5**, 1998, pp 137-151.
7. Komorowski, J. P., Simpson, D. L., and Gould, R. W., "A Technique for Rapid Impact Damage Detection with Implication for Composite Aircraft Structures", Composites, 1990, pp 169-173.
8. Gould, R. W., and Komorowski, J. P., "Detection of Impact Damage and Delaminations with Large Area Composite Inspection System, DAIS-500", NRC IAR LTR-ST-2031, October 1995.
9. Karpala, F., Willie, D., and Komorowski, J. P., "NDI Data Organization Methodology for Life Cycle Management", ASNT Spring Conference and Fifth Annual Research Symposium, Norfolk, Virginia, March 18-22, 1996.
10. Gould, R. W., and Komorowski, J. P. "Method for Preparing Solid Surfaces for Inspection", United States Patent, 5,569,342, Oct. 29, 1996
11. Petrin, C., Annis, C. A. and Vukelich, S. I., "A Recommended Methodology for Quantifying NDE/NDI Based on Aircraft Engine Experience", AGARD-LS-190, 1993.
12. Spencer, F., Borgonovi, G., Roach, D., Schurman, D., and Smith, R., "Reliability Assessment at Airline Inspection Facilities, Vol. I : A Generic Protocol for Inspection Reliability Experiment", DOT/FAA/CT-92/12, I, March 1993.

C-141 Spanwise Splice Advanced NDI Method (Probability of Detection Experiment Results)

Mr. Roy T. Mullis
Warner Robins Air Logistics Center (WR-ALC)
Technology and Industrial Support Directorate
Materials Analysis Team (TIEDM)
420-2nd Street, Suite 100
Robins Air Force Base, GA 31098-1640, USA

SUMMARY

Second-layer cracking of the lower inner-wing spanwise splice-joints was identified as the life-limiting structural feature of the C-141 aircraft. This cracking problem dictated the need for a new inspection process. The WR-ALC Materials Analysis Team (TIEDM) was tasked to develop a nondestructive inspection (NDI) procedure with a proven capability to detect 0.125 inch cracks in the splice-joint 2nd layer. TIEDM determined the best alternative inspection method, with potential to meet the 2nd layer inspection requirement, was an automated ultrasonic scanning technique. TIEDM contracted with SAIC/Ultra Image International for splice-joint inspection process development. SAIC subsequently designed a prototype ultrasonic scanning inspection system that met the C-141 requirements.

A Probability-of-Detection (PoD) Experiment was designed and conducted to formally quantify the inspection reliability of the prototype process. The PoD study simulated on-aircraft inspection conditions as close as possible by utilizing actual C-141 components for test specimens. A total of 16 test specimens were subjected to artificial cyclic loading to produce a statistically desirable fatigue crack population. The cracked specimens were subsequently characterized and documented, then assembled per established Air Force maintenance requirements. Fourteen inspectors with various training and experience backgrounds participated in the PoD experiment at WR-ALC. The experiment results show the new procedure has a 90% crack detection threshold of 0.073 inch. This data will allow the C-141 structural managers to confidently implement the new NDI procedure and establish future inspection intervals and requirements. In addition to providing reliability data, the PoD experiment also provided an information base on the procedural and human variables which most effect procedure results. This information will be used to make procedure enhancements to further improve the system reliability.

BACKGROUND

Inspection Area Description

The C-141 inner wing lower surface is constructed of 11 wing panels attached at spanwise splice-joints with 0.250 inch to 0.375 inch diameter taper-lok fasteners. A corrosion inhibiting sealant is applied in the faying surface of the panel-to-panel joints. Approximate splice-joint thickness (two layers) ranges from 0.275 inch to a maximum of 0.825 inch. The C-141 splice-joint configuration is detailed in Figure 1. Cracks initiate at the forward and/or aft side of the splice-joint inner tab fastener holes. Cracks that initiate on the forward side of a splice-joint fastener hole propagate until the edge distance is traversed and the ligament is severed. Cracks that initiate on the aft side of a splice-joint fastener hole propagate through the tab and the radius until surface breaking, as shown in Figure 1.

Paper presented at the RTO AVT Workshop on "Airframe Inspection Reliability under Field/Depot Conditions", held in Brussels, Belgium, 13-14 May 1998, and published in RTO MP-10.

Inspection System Description

An SAIC developed procedure was prototyped and validated through a series of laboratory and on-aircraft demonstrations. This prototype system was used to perform the PoD experiment. The inspection system utilizes an Ultra Image IV imaging system for all ultrasonic parameter and scanner motion control. Two shear-wave transducers, as shown in Figure 1, are employed to penetrate through the first-layer tab, the sealant bondline, and then, into the second-layer tab of the splice-joint. A two-axis scanner attached to the aircraft wing moves the transducers over a programmed inspection area. Precise gating of the ultrasonic signals in the second-layer tab provides the operator with an archived, easily interpreted C-scan image of the splice-joint fastener holes and associated defects.

1.0 PROBABILITY OF DETECTION EXPERIMENT DESIGN

An experiment was designed to determine the field achievable inspection reliability of the SAIC developed 2nd layer inspection procedure. Experiment design and execution were based on the guidelines of Reference 1. The generic protocols of Reference 1 were modified to specifically accommodate the C-141 splice-joint process. To obtain a better understanding of factors that could affect the reliability of the procedure the experiment was divided into two major phases: a laboratory validation phase and a field implementation phase.

1.1 LABORATORY VALIDATION

The intent of the laboratory validation phase was to characterize the impact of procedural variables on detection, as well as on the quality of signal. Five major procedure variables were identified and included in the laboratory validation experiment as factors to be studied in a fractional factorial experiment. Table 1 identifies the five variables and the assignment of the high and low levels of each that were used. These levels were chosen to reflect reasonable variation of each variable during actual inspections. The first three variables; timebase delay, depth velocity, and receiver gain are determined by the inspector during the calibration sequence of the procedure. The last two variables, scanner skew and probe pressure, reflect the major procedural aspects associated with the physical placement of the scanner with respect to the inspection sites.

A total of 17 experimental runs were performed during the laboratory validation experiment. Each experimental run is defined as a scan of all six specimens included in a block. Two blocks of specimens contained approximately the same distribution of cracks, as well as a similar distribution of test specimen thickness. Each combination of high and low variable levels was included an equal number of times and balanced in a manner to make the main effects of each variable clearly identifiable.

The laboratory validation conditions differed from field experiment conditions primarily in that the scanner was fixed and inspections were accomplished in an inverted position. The same lead operator conducted all laboratory experiment runs. The operator performed all calibration sequences per the written procedures. The time base delay, depth velocity, and receiver gain values as determined during calibration were used as the nominal values of the experiment. All scan image and ultrasonic data from the laboratory experiment were automatically saved. All crack / no-crack calls were made per the reporting criteria of the written procedure and recorded manually by the operator.

1.2 FIELD IMPLEMENTATION

The field implementation experiment was designed to emulate on-aircraft inspection conditions as close as practicable. Fourteen inspectors participated in the field experiment, each performing evaluation on separate days. All field experiment operations were monitored by government and contractor personnel. The monitors conducted in-briefings, collected data, conducted exit briefings and ensured that all operations were performed consistently from inspector to inspector.

An extruded aluminum framework was erected and used as the experiment inspection platform. The test frame design incorporated an actual full length C-141 wing panel to simulate the lower inner wing surface. The wing panel served as the mounting surface for the scanner as it would during an aircraft inspection. Eight test specimens that best represented the desired crack distribution were used during the field experiment. The test specimens were mounted in the inspection platform at the proper height to simulate overhead work on a typical C-141 wing stand. Test specimens were butted against the wing panel and mounted end-to-end to simulate one entire splice-joint length.

Inspectors were provided with the C-141 splice-joint 2nd layer procedure which detailed calibration and inspection requirements. All the equipment necessary to conduct the procedure was also provided. Using the procedures and equipment provided the inspectors performed a complete inspection of eight mounted test specimens. The inspection sequence started at one end of the frame and progressed to the other end through a series of scanner moves. All inspectors inspected the same eight test specimens in the same order. Following each scanning sequence the inspector evaluated the generated C-scan and manually recorded all inspection findings (crack / no-crack) on an inspection work sheet. The C-scan was then saved to hard disk in the inspectors designated directory. When the entire inspection was complete the inspector's findings were collected, data was backed-up and the equipment was readied for the next inspector.

1.2.1 Participants

The 14 experiment participants were minimum MIL-STD-410 Level II (or equivalent) qualified in the ultrasonic inspection method, but represented a vast cross-section of training and experience in the automated ultrasonic imaging technique used during the experiment. Table 2 summarizes the participants prior NDI experience. This cross-section of inspectors was purposely chosen to determine if and how these factors affected inspection results.

Three distinct inspector pools received different levels of training based on prior experience levels. The "expert" pool, also called the UI inspectors, included six operators with prior formal training and experience with the Ultra Image IV Ultrasonic System. The expert operators received one day of C-141 splice-joint procedural specific training. The "intermediate" inspector pool, also called the TI inspectors, included three technicians familiar with operation of the Ultra Image IV system, but with little or no production experience. The intermediate operators received one-week of equipment operation refresher training , including procedural specific training. The "novice" inspector pool, also called the LJ inspectors, included five technicians with no prior experience with the Ultra Image IV (or similar) equipment. The novice operators received two-weeks of extensive equipment operation training, including procedural specific training.

1.2.2 Common Data Set Evaluation

In addition to performing scans and evaluating data gathered during their own inspection, ten of the field experiment participants evaluated a

common data set that consisted of inspection images gathered during the nominal runs of the laboratory validation phase. The scanning and evaluation of test specimens during the field experiment represented a combination of procedural and data interpretation factors. Whereas, evaluation of the common data set allowed the data interpretation component of the inspection to be separated and characterized. Each of the ten inspectors was provided with the same common data set of splice-joint C-scan images and asked to evaluate the data per the C-141 splice-joint 2nd layer procedure acceptance criteria. All inspection findings (crack / no-crack) were recorded manually by the inspectors.

1.3 TEST SPECIMEN DESIGN AND CHARACTERIZATION

The 12 test specimens used in the PoD experiment were actual C-141 spanwise joint segments. The entire lower inner wing spanwise joint thickness range is represented by the segments. The segments consist of two separate pieces of adjacent wing panels fastened together at the splice-joint. Each segment is approximately 40 inches in length, and contain 22 to 38 fastener sites.. A typical segment configuration is shown in Figure 2. Initial characterization of the segments by SAIC included thickness measurements, hole measurements and immersion ultrasonic inspection.

Each segment piece containing the inner tab was cyclic fatigue tested to obtain an acceptable distribution of 2nd layer cracks. The approach to obtaining controlled crack growth was to sequentially initiate cracks, smallest to largest, with 0.020 inch starter notches. The notches were cut into predetermined fastener sites and cracks grown to the desired lengths. A total of 62 crack sites were initially verified by optical inspections. Crack directions were distributed in the forward (33) and aft (29) directions. The cracked holes were oversized to remove starter notches. Final crack length measurements, following hole oversizing, are summarized in Table 3. Following hole oversizing and final optical crack characterization, the segments were assembled with taper-lok fasteners and faying surface sealant. All segment assembly operations were performed in accordance with Air Force maintenance standards. Once assembled, the segments were painted with an approved C-141 exterior coating system. Final characterization was performed on each panel with ultrasonic immersion testing.

2.0 DATA ANALYSIS

Data analysis was performed independently by Sandia National Laboratories (SNL) under sub-contract to SAIC/Ultra Image International, Inc. Following is a summary of the SNL report.

2.1 ANALYSIS OF LABORATORY DATA

It should be noted that the data taken during the laboratory validation was not a blind experiment as the laboratory operator was familiar with the test specimens and crack locations. In order to provide unbiased results, only the data collected during the field validation experiment will be used to estimate the overall system reliability. Nevertheless, the laboratory data provided valuable information on the impact of procedural variables on inspection reliability.

2.1.1 Analysis of Laboratory Calls

The laboratory validation portion of the experiment resulted in each of the 12 test specimens being inspected nine times. That is, a nominal scan, and eight additional scans with set-up parameters varied in a controlled manner. The twelve test specimens provided 356 inspection sites (fastener holes) including 58 cracked sites.

Of the 356 sites 341, or 95.8%, were consistently called or not called by both transducers in all nine runs. Another six, or 1.7%, were consistently called by one transducer and consistently not called by the other. That leaves nine sites (2.5 %) that were sometimes called and sometimes not called by at least one of the transducers. The implication is that changes in the various inspection set-up parameters can impact the resultant signal characteristics and, in turn, signal interpretation enough to affect inspection reliability. Various aspects of the signals in relation to the setup parameters are analyzed in the next section.

2.1.2 Laboratory Signal Analysis

The C-scan image in Figure 3 shows the signals of two adjacent fastener holes from two separate laboratory runs. Both runs were performed with the same transducer. The hole pair on the bottom of Figure 3 show typical fastener and crack indication images. The hole on the left side contains a crack as is indicated by the crack indication or "lobe" to the bottom left of the fastener hole image. No crack indications are present at the hole to the right. The hole pair on the top show an image of the same two fasteners following parameter changes. The fastener hole and crack indication images in the top hole pair show large variations in signal area and amplitude attributable to the set-up differences. Signal characteristics were further analyzed to determine the affects of individual parameters. The analysis was accomplished by measuring the fastener hole signal amplitude and crack signal area and amplitude from various laboratory validation C-scan images. The signal area and amplitude data were then studied in relation to the controlled parameters included in the experimental design: time base delay, depth velocity, gain, skew, and probe pressure. In addition, the interaction between the timebase delay and depth velocity was included in the analysis.

The total estimated effect (change in signal response between the high and low levels) of each parameter is shown in Table 4. The effects are expressed as a percentage of the nominal run level. The entries in Table 4 indicate that the depth velocity, either by itself or in conjunction with the time delay, is a significant factor in explaining variations of the signal characteristics. The signal variation associated with depth velocity comes as a direct consequence of the differences in operator measurements of the transducer angle during procedure calibration. Similarly, gain was also indicated as a significant factor for most of the responses. The effect of gain was in the direction expected. That is, the higher the gain the "stronger" the signal. In the three cases where probe pressure was indicated as significant, an increase in the probe pressure resulted in "weaker" signals. Manually engagement of the transducer holder appeared to establish better contact than did engaging the holder remotely. Scanner skew, at ± 0.250 inch, did not appear to be a major contributor to signal differences in the laboratory experiment. The results of these analysis pinpoint sources of inspection variability and error. With this knowledge the system developers and users can make procedure and training changes accordingly.

2.2 ANALYSIS OF FIELD DATA

2.2.1 Calibration and Setup Parameters

It was detailed in the previous section that the depth velocity variations significantly contribute to signal variability. This velocity value is based directly on the measurement of the transducer angle by the inspector during calibration sequence. Table 5 shows the variation in transducer angles measured and recorded during calibration in the field inspections. The variation in angle measurements reflected in Table 5 indicate that the two degree variation of

transducer angle used in the laboratory experiment reasonably reflects the uncertainty in the procedure. Table 5 also shows the range of gain and timebase delay values used during the field experiment. However, for the gain and time-base delay, the field data indicates that the range studied in the laboratory phase probably does not accurately reflect the real-case uncertainty. Again, this data is valuable to the users and developers for making procedure improvements to lessen these degrees of uncertainty.

2.2.2 Errors in Reporting Flaw Location

During the field experiment, inspectors manually recorded call / no-call data onto inspection work sheets by indicating whether a crack was present at each site. A database was established using each inspectors' work sheet. While entering the field data into the database, it was noted that some of the missed cracks had false calls made at the immediately adjacent site. Upon review of the saved scan files it was obvious in many cases that missed calls were due to hole misidentification. In the data analysis that follows curves fit to the data as delivered are compared to those fit to the data corrected for misidentifications.

2.2.3 Probability of Detection Curves

Figure 4 shows four different curves fit to the full set of field experiment data. The curve farthest to the right at a probability of detection of 0.90 is the traditional two-parameter probit curve. The high detection portion of this curve is extended to the right because of several large crack misses attributable to hole misidentification by one or more of the inspectors. A four-parameter curve, as proposed by SNL in their analysis, models the larger cracks being missed at an approximate rate of 3.5 percent, due to factors other than crack length. This is further proven by correcting the data for known misidentifications and rerunning the curve fits. The two-parameter and four-parameter models fit to the "aligned" data are also included in Figure 4. The aligned curves show that most of the 3.5 percent estimate was, in fact, due to misidentification. The four-parameter fit to the aligned data yields a 50% detection crack length of 0.038 inch, and a 90% detection crack length of 0.073 inch. The overall false-call rate of the field experiment was 0.091 as estimated from the non-cracked fastener hole population.

2.2.4 Individual Inspector Data

Individual inspector probability of detection data is summarized in Table 6. The PoD data provides no indication that the three populations of inspectors performed any differently. This is clearly shown by the Table 6 data for Inspectors 11, 12, and 13, all from separate inspector pools. The inspectors have almost identical 50% crack lengths, only a 0.010 inch spread for the 90% crack lengths, and false call rates all under 3.0 percent. The wide range of 90% crack lengths (0.053 to 0.166) and false-call rate (0.018 to 0.401) shown in Table 6 indicates more in-depth inspector training may be necessary to bring these ranges to a more consistent level.

2.2.5 False-call Rate

As previously noted, the false-call rate among the fourteen inspectors ranged from 0.018 to 0.401. The inspector with the 0.401 false-call rate also had the lowest 50% crack detection length and near the lowest 90% crack detection length. The high false-call rate coupled with the low crack detection lengths suggests this inspector was ultra-conservative with calls made on the C-scan data.

Removing this anomalous rate from the field data reduces the false-call rate spread from 0.018 to 0.155 and the overall false-call average

down to 0.067. This rate is more consistent with the 0.073 false-call rate obtained from the common data set discussed next.

2.3 ANALYSIS OF COMMON DATA SET

Ten inspectors evaluated a common data set constructed of images gathered at nominal settings during the laboratory experiment. The common data set contained 356 images in 30 different data files. The inspectors were asked to evaluate the images using the same criteria as used during the experiment. By studying the differences in the calls made on this common data set, we can separate issues of set-up and calibration from those of the decision process associated with signal interpretation.

Figure 5 shows the four-parameter curve fit to the common data set. The 90% crack detection size is 0.076 inch, consistent with 0.073 inch length determined from the field experiment data. The average false-call rate of the ten field inspectors on the common data set was 0.073. This rate is close enough to the field experiment rate (0.091) to indicate that the variation of signals from multiple inspections is not the major contributor to the false-call rate. Further analysis on the common data set false-calls using a collective decision model resulted in rate of 0.035. This analysis reinforces that the decision process involved with signal interpretation accounts for at least half of the observed false call rate. Overall, the common data analysis suggests that more intensive training is required in this area to lower the variation in signal interpretation from inspector to inspector.

3.0 SUMMARY AND DISCUSSION

The experiment proved very successful. The results provided not only quantitative inspection sensitivity data, but also valuable details on procedural variability and inconsistencies. The experiment shows the 90 percent detection capability of the new splice-joint inspection process is a 0.073 inch 2nd layer crack. This is far more sensitive than the original 0.125 inch detection capability requested by the C-141 management. The 0.073 inch detection threshold would allow the establishment of a 5-year inspection interval for the splice-joint area. This would relieve base-level personnel from this inspection burden and save the Air Force approximately $50 million in C-141 inspection and repair costs over each 5-year inspection period.

The 9.1 percent false-call rate is higher than expected, but with the data provided by the field and laboratory experiments, it is anticipated that the rate can easily be lowered to $\leq$ 4.0 percent. Six of the field experiment participants had false-call rates 4.1 percent or less. The laboratory phase of the experiment provided an information base on the effects of procedural and human variables on the procedure reliability. With knowledge of this information base, a procedure and training syllabus review will be conducted. The emphasis of the review and any subsequent procedure and training revisions will focus on areas pinpointed by the experiment data (angle measurement, gain settings, delay settings, hole misidentification, and signal interpretation). Favorable changes in these areas will help lower the false-call rate and also increase the inspection reliability.

4.0 REFERENCES

1. Spencer, F.W. and Schurman, D. L., "Reliability Assessment at Airline Inspection Facilities, Volume I: A Generic Protocol for Inspection Reliability Experiments," DOT/FAACT-92/12,I, March 1993.

Table 1 Experimental Variables and levels

Variable		Low level (-1)	High level (1)	Nominal level (0)
1. time base delay		nominal - 0.005	nominal + 0.005	determined in calibration
	used	0.350-ch 1 0.355-ch 2	0.360-ch 1 0.365-ch 2	0.355-ch 1 0.360-ch 2
2. depth velocity		table value for probe angle - 1 °	table value for probe angle + 1 °	tabled value determined from probe angle
	used	85,400 in/sec	88,500 in/sec	87,000 in/sec
3. receiver gain		nominal - 0.6 dB	nominal + 0.6 dB	as determined at time of calibration
	used	35.60 dB	36.80 dB	36.20 dB
4. scanner skew		0.250 inch left	0.250 inch right	centered
5. probe pressure		pressure off	nominal	arbitrary
	used	0 - 1 lbs. indicated	16 lbs indicated	16 lbs indicated on dial

Table 2 Inspectors Background

	Inspector													
	LJ01	LJ02	LJ03	LJ04	LJ05	UI01	UI02	UI03	UI04	UI05	UI06	TI01	TI02	TI03
Years Experience in:														
NDT	10	26	19	16	25	10	15	15	22	7	11	20	20	25
UT	9	24	19	16	25	10	15	15	22	7	10	18	20	25
computerized methods	0	0	0	0	0	6	15	4	2	0.2	2	2	1	1
Aircraft	9	17	17	16	25	10	4	4	22	11	11	25	17	25
C-141	9	17	8	0	24	6	0	0	0	0	0	25	17	25
2nd layer	3	1	0	0	0	0	4	3	2	0.2	2			0.5

Table 3 Specimen and Crack Distribution

	Test Sections	Inspection Sites	Crack length (inch)				
			< 0.050	0.05 to 0.10	0.10 to 0.15	0.15 to 0.20	> 0.20
Laboratory phase	12	356	16	18	7	9	4
Field phase	8	238	15	16	6	7	4

Table 4 Effects of Experimental Factors on Selected Signals

	Experimental Factors					
Response	time delay	depth velocity	gain	skew	probe pressure	time delay * depth vel.
Inner Transducer						
Average signal strength	in interaction	in interaction	7.8		1.2	2.9
Area of flaw signal		11.2	17.0			
Average signal strength of flaw area	in interaction	in interaction	7.1			15.4
Outer Transducer						
Average signal strength		4.4	9.1	2.9	6.3	
Area of flaw signal	in interaction	in interaction				19.8
Average signal strength of flaw area		4.7	9.9		4.6	

Table 5 Calibration values for 14 field inspections

	angle*			gain			time	base	delay
	45°	46°	47°	median	min	max	median	min	max
Transducer 1	5	2	7	33.9	17.6	38	0.342	.298	.421
Transducer 2	9	3	2	34	17.2	39.8	.373	.018	.435

* tabulated values are the number of inspectors measuring the given angle

Table 6 PoD curve points for individual inspectors

Curve #	Inspector class	*false call rate	50 % crack length (inch)	90 % crack length (inch)
1	LJ	0.111	0.036	0.053
2	TI	0.122	0.036	0.053
3	LJ	0.401	0.025	0.057
4	TI	0.036	0.034	0.060
5	LJ	0.053	0.051	0.060
6	LJ	0.081	0.033	0.061
7	UI	0.155	0.040	0.067
8	UI	0.094	0.042	0.070
9	UI	0.041	0.042	0.070
10	UI	0.076	0.033	0.072
11	UI	0.029	0.043	0.076
12	LJ	0.018	0.043	0.078
13	TI	0.018	0.042	0.086
14	UI	0.034	0.053	0.166

* estimated from non cracked population - (.) values estimated from maximum likelihood.

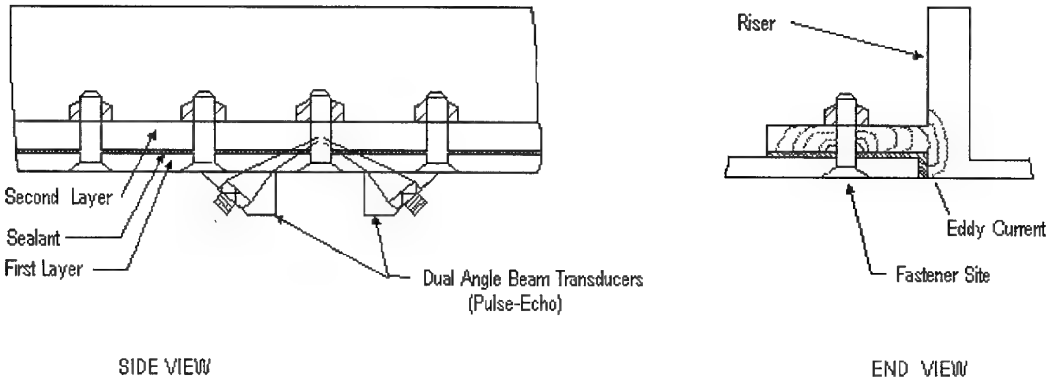


Figure 1 Splice-Joint Configuration.

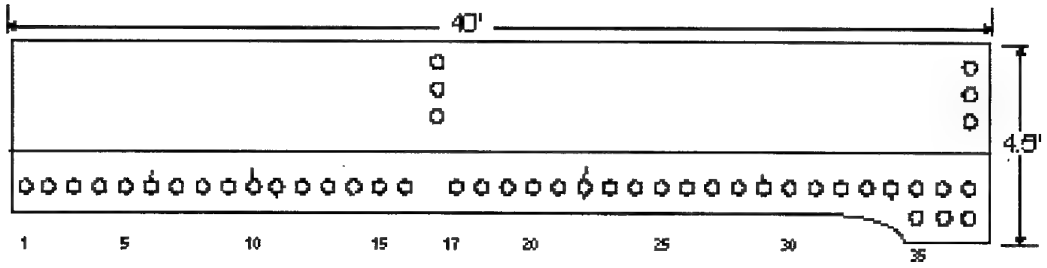


Figure 2 Typical POD Test Specimen. Second Layer Only Shown.

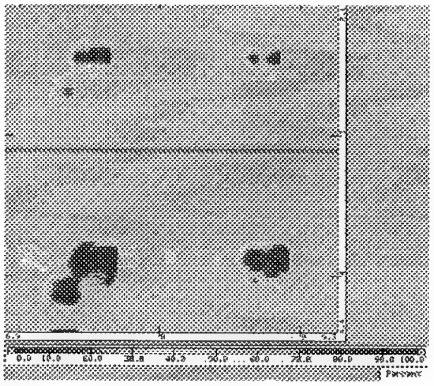


Figure 3 Fastener Site and Crack Signal Examples.

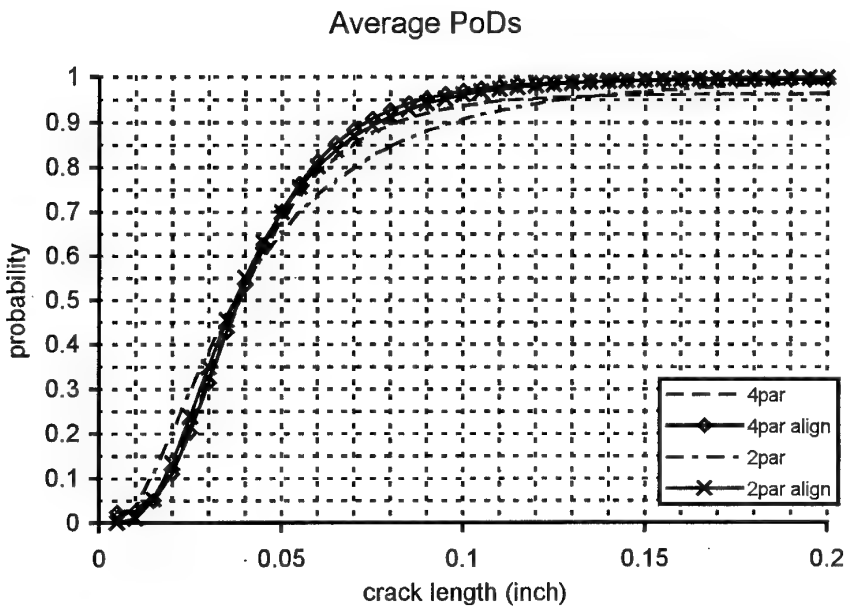


Figure 4 PoD Curves fit to Field Experiment Data Set

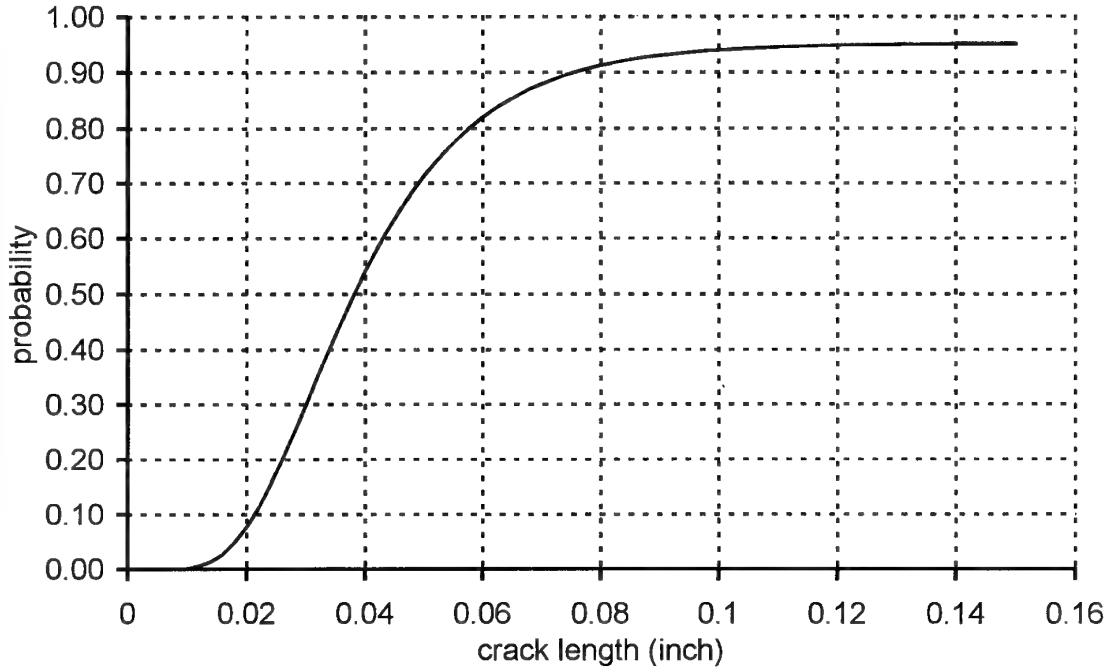


Figure 5 Average PoD Curve From the Common Data Set

A NEW APPROACH FOR RELIABLE INSPECTION IN RADIOGRAPHY OF TURBINE BLADES

Asst. Prof. Dr. Hacer AYGÜN
Izmir Institute of Technology, Department of Mechanical Engineering,
35230, Cankaya, Izmir-TURKEY

Prof. Dr. Ekrem SELÇUK
Middle East Technical University, Department of Metallurgical and Materials Engineering
06531, Ankara-TURKEY

1. SUMMARY

Some extra ordinary patterns have been observed on the radiograps for Ni-base gas turbine blades, which cannot be correlated to any kind of flaws (or other density differences) in the material and which commonly are called "mottling".

In this study, reasons of the mottling have been investigated by radioscopic and radiographic methods. The major mottling indications are caused by Laue diffraction of the tungsten target material characteristic K_α and K_β radiations from the thin edge of the specimens, <200> solidification direction and FCC diffraction plane.

To prevent mottling on the radiographs, a new double-slit system and a new NDT-slit system have been developed by the authors. The results show that new developed NDT-system is the best technique for preventing mottling indications on the radiographs of the directionally solidified Ni-base gas turbine blades.

2. INTRODUCTION

The raising request for high performance alloys in the gas turbine technology has led to the development of directionally solidified and single crystal turbine blades, which need individual inspection by radiography or real-time radioscopy in order to assure reliable performance under extreme thermal and mechanical conditions.

A special from of scattering (mottling) caused by X-ray diffraction is encountered occasionally in radiographs of directionally solidified Ni-base turbine blades. The mottling appearing on radiographs usually consists of a roughly straight dark line along the axis of the solidification direction and FCC diffraction planes. The radiographic appearance of this type of scattering is mottled and may be confused with the mottled appearance sometimes produced by porosity, cracks or segregation (1).

Mottling, until now, has been explained by different effects such as X-ray diffraction from grain boundry, X-ray scattering from porosity, segregation and inhomogenities (2, 3, 4).

However there is a general agreement to consider it a major problem in radiography of strongly textured materials, coarse grain, single crystal component and face-centered cubic structural materials such as aluminum, stainless steel and nickel base alloys (3, 5, 6, 7).

A relatively large crystal or grain in relatively thin specimens may in some cases reflect an appreciable portion of the X-ray energy falling on the specimen, much as if it were a small mirror as shown schematically in Fig.1. The diffracted beam strikes the film and will resulting some dark lines (mottling) on the film.

This effect is not observed in most industrial radiography, because most specimens are composed of a multitude of very minute crystals or grains, variously oriented; hence, scatter by diffraction is essentially uniform over the film area. However in the radiographs of directionally solidified or single crystal specimens mostly there are mottling indications.

A lot of different proposals are given to influence and prevent mottling: The mottling indications caused by X-ray diffraction can be reduced, and in some cases eliminated, by raising tube voltage and by using lead foil screens on front of the tube (2, 5, 8, 9). Another approach used to distinguish mottling and defects by making two succesive radiographs, with specimen rotated slightly (1-5 degrees) between exposures, about an axis perpendicular to the central beam. Therefore, a pattern caused by porosity or segregation will only change slightly; however, one caused by diffraction will show a marked change (2).

Recently, the problem of mottling has been treated by image re-constraction methods, where two radiographs taken under different angles are compered for coincidence of the diffraction patterns (10). In this system grey level of mottling indications on the two radiographs are corrected by image process, then re-constracted again.

In this research, radiographic and real-time radioscopic testings of directionally solidified Ni-base gas turbine blades has been performed to establish the reasons of the mottling and prevent mottling by recommended methods. Additionally, the aim of this study is also develop new techniques to eliminate mottling completely. For this reason, a new double-slit system and a new NDT-slit system have been developed by the authors and applied to the radiographic testing succesively.

Paper presented at the RTO AVT Workshop on "Airframe Inspection Reliability under Field/Depot Conditions", held in Brussels, Belgium, 13-14 May 1998, and published in RTO MP-10.

3. MATERIALS AND METHODS

Mottling indications have been investigated on stationary gas turbine blades of <002> directionally solidified Ni-based alloys (IN 792 DS and IN 6203 DS) of 15 cm blade length. Blades contain a face-centered-cubic nickel matrix strengthened by app. 40% volume fraction of coherent intermetallic γ-Ni_3 (Al,Ti) and small volumes (<1%) of various carbides, borides and carbosulphides.

Radioscopic examinations of the blades were performed by using ANDREX-MX4 microfocus equipment with 20-30 μm focal spot size of tungsten target X-ray tube, image intensifier, sample handling system (manipulator), image process system and image documentation system such as video-printer and video-band. Conventional film-type radiography has also been used for documentation of the images. The experimental set-up is shown schematically in Fig.2.

In order to demonstrate the occurance of Laue diffraction, a simple pin-hole collimator of 1 mm diameter was installed at 60 mm distance from the X-ray focus together with the sample. The transmission Laue diffractogram was taken from the thin part of the turbine blade at 160 keV tube voltage and 1.0 mA. Diffraction angle of mottling which coincides with FCC diffraction planes were calculated on the radiographs with known focus-specimen and focus-film distance.

To study the diffraction patterns of the complete oriented turbine blade by mottling on radiographs, the blade was surrounded by a lead mask of 1mm. This arrangement has the advantage that the mottling patterns can be observed in this area of the lead mask shadow, where it is not superimposed to the radiographic image of the sample.

Prevention of mottling on the radiographs, previously recommended methods were used such as turning the specimen, increasing tube voltage and using filters.

Suppressing scattered radiations has been performed by a new developed double-slit system. The double-slit system arrangement is shown schematically in Fig.3, which employs two parallel slits at a distance "c". The limiting angle of diffraction can be chosen by variation of the slit widths "c_1" and "c_2". The sample and the film provide the possibility of adding the individual radiographic slit images the film. For experimental simplicity, the same speed for the sample and the film by a common support on one manipulator was used.

Radiographic testing of turbine blades with new developed double-slit system takes very long exposure times and contrast of the radiographs is not enough for evaluation. Therefore, a new NDT-slit system has been developed by the authors. This system is composed of lamels as 8x100 mm lead plate of 0.2 mm thick stacked on top of each other and laminated with polyethylene sheets and assembled in an aluminum frame. Lead lamels were seperated with 0.3 mm polyethylene that has very low attenuation coefficient and permits all radiation. Polyethylene with 0.3 mm thick gives the intensity ratio of outcoming to incoming intensity ratio (I_x / I_o) as 0.994 that is nearly one.

Slit period is 0.5 mm with 0.2 mm lead plate and 0.3 mm polyethylene. Maximum opening angle of slit was 4.3° that is less than tungsten characteristic diffraction angle (6°). A beveling angle of 9.5° ensures that the X-ray tube focal spot sees a constant slit width during scan, even if only a single slit is used. During exposure, the NDT-slits moved chaotically between the sample and the film with help of springs at four edges.

4. RESULTS AND DISCUSSION

4.1. The Causes of Mottling

Overcoming of mottling problems is possible with determining the causes . An experiment was carried out with Scherrer method to determine lower limit of diffraction angles of the IN 792 DS blade. For a given Cu-K_α wavelengths with Ni-filter, the diffraction planes and basic net-plane spacings "d" have been calculated from the Scherrer diagram, given in Fig.4.

From the Bragg's relation;

$$\lambda = 2d_{hkl} \sin\theta = 12.4/E \qquad \text{(Eq.1)}$$

Where

d: distance between lattice planes, A°
θ: diffraction angle,°
λ: wavelength of x-ray, A°
E: X-ray energy, keV
hkl: Miller indices.

The basic netplane spacings "d" have been found as:

$d_1 = 1.77$ A°
$d_2 = 2.02$ A°

which correspond to the well known respective spacings d_{200} and d_{111} of nickel with cubic cell dimension of a=3.52 A°.

The typical appearance of mottling indications from the sample is shown on the radiograph of Fig.5. The mottling indications can be understood as the superposition of many single crystals rotational Laue diffraction patterns, with the vertical solidification direction of 200-direction. The direction of the black streaks can be identified as running perpendicular to the lattice plane direction of a FCC lattice as demonstrated in Fig.6, where the streaks are re-drawn. The directions of streaks are found to belong to the netplanes of low Miller indices, such as <200>, <220>, <420> and <002>.

As it can be seen on the radiograph (Fig.5), mottling indications are resulting from the diffracted beams from FCC diffraction planes like (111) and (420) and mostly (200) of solidification direction.

In order to demonstrate the occurance of Laue diffraction more clearly with using a simple pin-hole collimator, Laue transmission diffractogram was taken from the thin part of the turbine blade (Fig.7). The strongest doubled spots on the diffractogram are identified to be the <002> reflections of 58 keV and 67.3 keV which are K_α and K_β characteristic tungsten radiations. Due to the low energy level of K_α and K_β lines can not be absorbed by the specimen and according to Bragg's diffraction relation (Eq.1), an appreciable portion of X-ray with low energy "reflects" like a small mirror at relatively thin sections.

In normal radiograph, the surroundings of the sample are dark and nothing can be seen. Therefore, to study the diffraction patterns, the blade was surrounded by a lead mask of 1 mm. The radiograph taken with a lead mask at 160 keV with mottling indications is shown in Fig.8. Mottling caused by characteristic radiations diffraction of K_α and K_β of tungsten can be observed inside and outside the sample. Since the sample of film distance and the distance between the mottling indications and the edge image are known, a scattering angle of 6° was calculated from Bragg's formula. Also 6° (2θ) is the same angle with K_α characteristic line of tungsten and also as that of the corresponding Laue intensive spots.

4.2. Prevention of Mottling

Previous researchers' recommendations, such as turning the specimen, increasing tube voltage and using filter had to be tested.

First, a radiograph with mottling indications was taken at the reference point (Fig.9.a), then at the same exposure conditions the sample was turned to the right with 2 and 5 degrees (Fig.9.b and c). As can be seen the Figures 9.b and c, the white images of mottling indications only changed the place, but not disappeared. Still one can confuse the mottling with defect indications in the radioscopic images.

Eliminating mottling by turning the specimen has been tested at different angles (from 0 to 42° with 2° augmentation) and found that the sequence of mottling repeating is app. 8°, which it may be the angle of orientation of one grain to the others. Because, if mottling indications come from scattering of one certain diffraction plane of one grain, after turning the specimen with 8°, the scattering comes from the same diffraction plane of the grain. Consequently, it can be said that during directional solidification, the grains are oriented app. 8° to the adjacent grains.

Results of this testing showed that turning the specimen is not effective way in eliminating the mottling.

With increasing tube voltage prevention of mottling indications was investigated on the same sample. An image was taken at 123 keV and 2mA (Fig.10.a), then at the same tube current the voltage was increased to 130 keV (Fig.10.b) and 155 keV (Fig.10.c). The increasing tube voltage to 155 keV decreased the mottling indication is some amount, but not completely eliminated. However increasing voltage decreases the contrast of the radiographs.

Another proposal is to use the filter. The energy spectrum of the tungsten target at 160 keV after transmission of the thin blade area without and with Cu-filter with various thicknesses has been analysed with a multi-channel analyser to show harden-up effect (Fig.11).

Remarkable hardening of characteristic lines can be observed with 2 mm Cu-filter in front of the tube. This effect is enforced by a 3 mm Cu-filter (lower spectrum). With 3 mm Cu-filter maximum intensity shifts to the high energy sides and characteristic K_α and K_β lines are decreased, and also soft radiation is absorbed by the filter.

The reduction of mottling indications by the application of Cu-filter is demonstrated by comparing the radiograph in Fig.12.a without filter and Fig.12.b with 3 mm Cu-filter. Using the filter decreases the amount of mottling. This effect can be easily understood by a relevant reduction of the characteristic radiation lines intensity and it supports the above statement which explains the mottling mainly is due to the diffraction of characteristic K_α and K_β radiations of tungsten target.

The results show that well-known or previously proposed methods are not efficient to eliminate mottling. Therefore a new double slit system and a new NDT-slit system were developed by the author to suppress the scattered radiation geometrically.

The working principle of the double-slit system is the absorption of the scattered radiation coming from the sample by help of lead screen on the slit places. Only mean beam or with very small divergence angle beam can pass through the slit opening and projected on the film. Mottling amount on the radiograph changes with the opening angles of the slit system. If the opening angle is smaller than the scattering angle, mottling can be completely eliminated. Therefore the opening angle was choosed 3.3°, which is below the limiting angle of K_α characteristic tungsten radiation.

The sample with mottling indications, shown in Fig.5, was scanned in horizontal (Fig.13) and vertical (Fig.14) directions by double-slit system with keeping the same exposure conditioned and positions. The radiographs are nearly free of scanning stripes. Comparisons of the scanned radiographs with and without double-slit systems show strongly reduction of mottling.

Scanning direction with slit-system is very important to prevent mottling. Diffractions at directions other than the scanning direction, pass through the slit opening and cause mottling. For example, in vertical scanning the scattered radiations in vertical direction are absorbed and diffraction only in horizontal direction is allowed to pass through the slit openings and in horizontal scanning is visa versa.

Another fact is that, scanning in one direction causes 1:1 dimensions with the sample, but due to sample-film distance there is always enlargement in other direction.

Clearly a double-slit system developed in this research is an effective method for removing scattered radiation, and thus decreasing the amount of mottling in the acquired image. However this method requires a long exposure time for sample scanning and it only absorbs the scattered radiation in one direction. These disadvantages can be improved in practice by employing medical multi-slit system (1). But it is not suitable for high energy radiographs due to thickness of lead lamellas. The thickness of lamellas are not big enough to absorb all scattered radiation .

A new NDT-slit system has been developed by the authors to obtain high quality images on the radiographs by detecting only primary radiation and eliminating all scattered radiation in a short exposure time, thus obtain the radiographs without mottling.

The new NDT-Slit system was designed for thick samples such as gas turbine blades.

Quality of radiographs taken by NDT-slit system depends on a number of physical parameters, such as slit thickness, slit period, opening angle and speed of chaotic movement. In the new NDT-slit system, the slit period is 0.5 mm, the opening angle is max. 4.3° and the bevelling angle is 9.5°.

Figure 15 shows a part of the blade with typical mottling indications. As can be seen on the radiograph, mottling indications are in vertical direction that are the same direction with the solidification direction. To absorb the scattered radiations from the edge and (200) solidification direction, first the NDT-slit system was placed vertically at the back of the blade (Fig.16). Mottling in the vertical direction was eliminated, however, there still remains some amount of mottling due to scattering from other directions.

The best method to eliminate completely mottling caused by scattering from FCC diffraction planes, is to use two NDT-slit systems as a cross system. Figure 17 shows the radiograph of the blade with cross NDT-slit system that absorbed all scattered radiations coming from all directions. Both NDT-slits were moved chaotically again during the exposure, and also the contrast and resolution of the radiographs are good enough to evaluate the defects of the blade.

5. CONLUSIONS

The results obtained in this study have allowed the following conclusions to be made:

1. Mottling indications on the radiographs of directionally solidified Ni-base gas turbine blades are caused mainly by X-ray diffraction from the columnar grains growing vertically upwards and FCC diffraction planes with low Miller indices such as <111>, <200>, <220>, <420> and <422>.

2. Tungsten characteristic K_α and K_β radiations with low energies (<60 keV) diffract at the thin side of the specimen with 6° diffraction angle and cause mottling.

3. Rotating the sample with small angles or increase tube voltage during exposure do not eliminate mottling.

4. It is possible to decrease the amount of mottling by using Cu-filter, but filters can not eliminate completely and cause reduction of contrast of the radiographs.

5. Diffraction from FCC diffraction planes can be suppressed by using double-slit system in which only mean beam can pass through the slit opening and projected on the film while the scattered radiations are absorbed by lead screens on the slit places.

6. A new NDT-slit system developed by the authors is the most effective system to absorb all scattered radiations and prevent mottling indications on the radiographs of directionally solidified Ni-base gas turbine blades.

6. REFERENCES

1. Aygün, H., "Investigation of Mottling in Radiography of Directionally Solidified Ni-base Gas Turbine Blades", Ph. D. Thesis, Metu, December 1996.

2. Metals Handbook, Vol. 11, Nondestructive Inspection and Quality Control, 8th Ed., (1976).

3. Saeki, Y., Sofue, T., Watanabe, T., Br. J. NDT, 19, (1988) 1.

4. Irie, J. of JSNDI, Vol. 12, No. 6, (1963).

5. Eastman Kodak Company, Radiography in Modern Industry, Rochester, New York 14650, 43, (1980).

6. Ryazantsev, V. I., Tolkachev, Y.R and Slavin, G. A., Weld. Prod., Vol. 33, (1986).

7. Watanabe, T., Saeki, Y., Mizukoshi; H., Br. J. NDT, 341; (1984) 9.

8. ASNT NDT Handbook, 2nd Ed., Vol. 3, Radiography and Radiation Testing, 211, (1985).

9. Watanabe, T., Saeki, Y., Br. J. NDT, 24, 181, (1982) 4.

10. Watanabe, T., Sofue, T., Saeki, Y., Muzokoshi, H., Welding Int., Vol. 2, 224, (1988) 3.

7. FIGURES

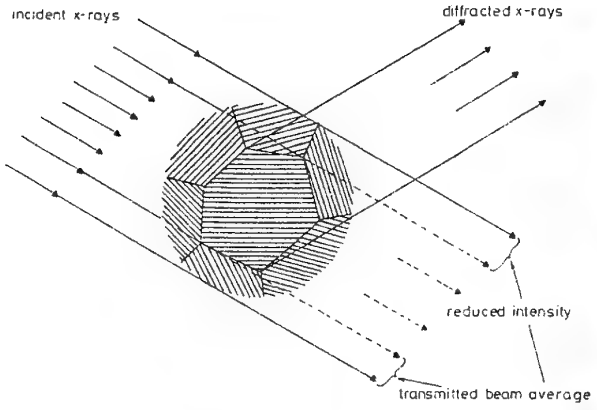


Fig. 1. Schematically explaining the reflection of some portion of the incoming X-ray from a relatively large grain.

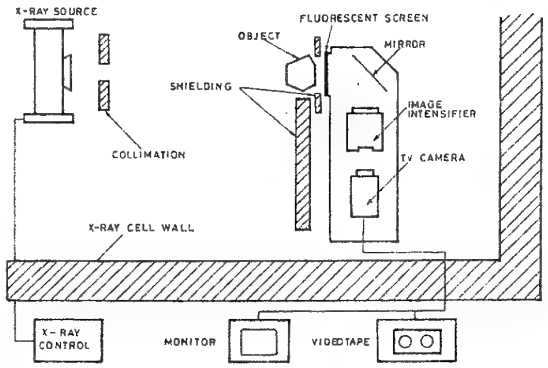


Fig. 2. Remote real-time radioscopic system.

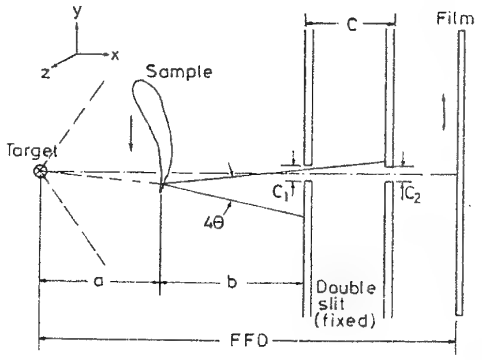


Fig. 3. Working principal of the double-slit system.

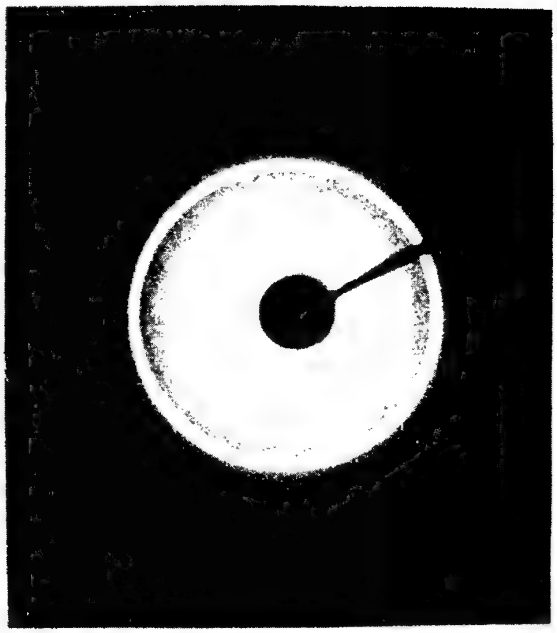


Fig. 4. Scherrer powder diagram, with the exposure conditions: λ (Cu-Kα): 1.542 A°, 40 kV, 30 mA, Ni-filter, t: 7 min, b (specimen-film distance): 25 mm.

Fig. 5. Radiographic image of mottling of the gas turbine blade, with the exposure conditions: 155 kV, 1.0 mA, D7 film, 0.02 Pb screens, FFD: 360 mm, a (focus-specimen distance): 140 mm.

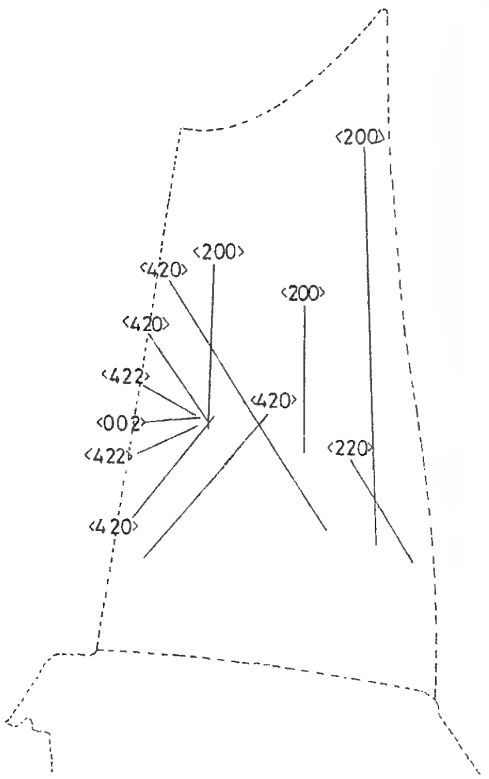


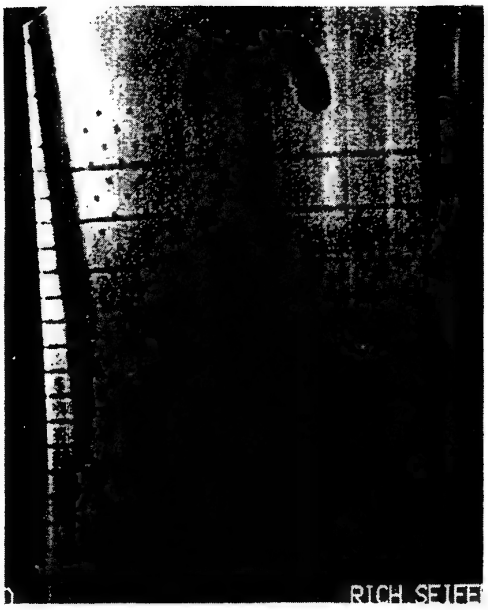
Fig. 6. Re-drawing of mottling indications of the gas turbine blade as FCC diffraction planes <420> and solidification direction <200>.

Fig. 7. Transmission Laue photograph, with the exposure conditions: 160 kV, 1.0 mA, t: 10 min, D7 film, 0.02 Pb screens, a: 120 mm, b: 60 mm, focus-collimator distance: 60 mm.

Fig. 8. Radiograph of the turbine blade, covering with 1 mm Pb-mask, 155 kV, 1.0 mA, t: 8.3 min, D5 film, 0.02 Pb screens, FFD: 700 mm, a: 568 mm.

(a)

(b)

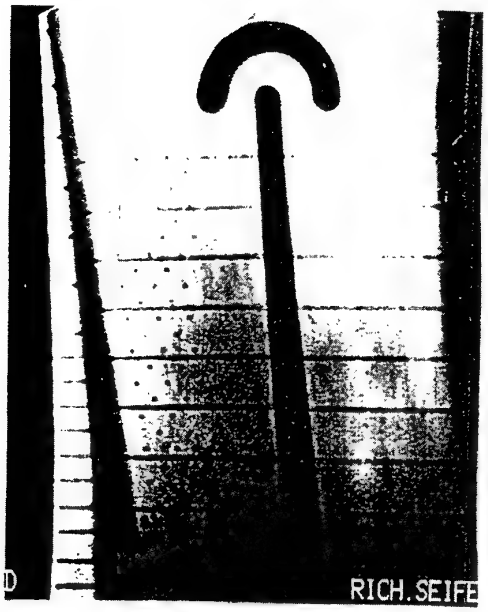


(a)

(c)

(b)

Fig. 9. Radioscopic images of mottling indications of IN 792 DS turbine blade, 90 kV, 4.0 mA, FFD: 700 mm, a: 625 mm;
(a) at reference point
(b) 2° turned to the right
(c) 5° turned to the right



(c)

Fig.10. Radioscopic images of mottling indications with increasing tube voltage; 2.0 mA, FFD: 700 mm, a: 585 mm.
(a) 123 kV
(b) 130 kV
(c) 155 kV

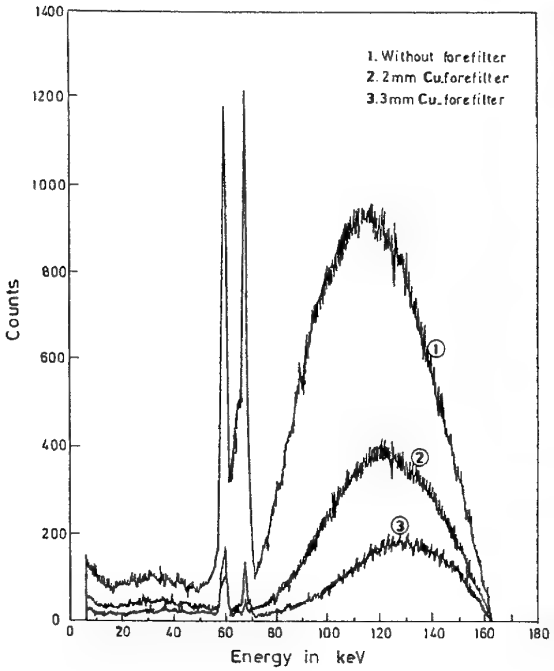
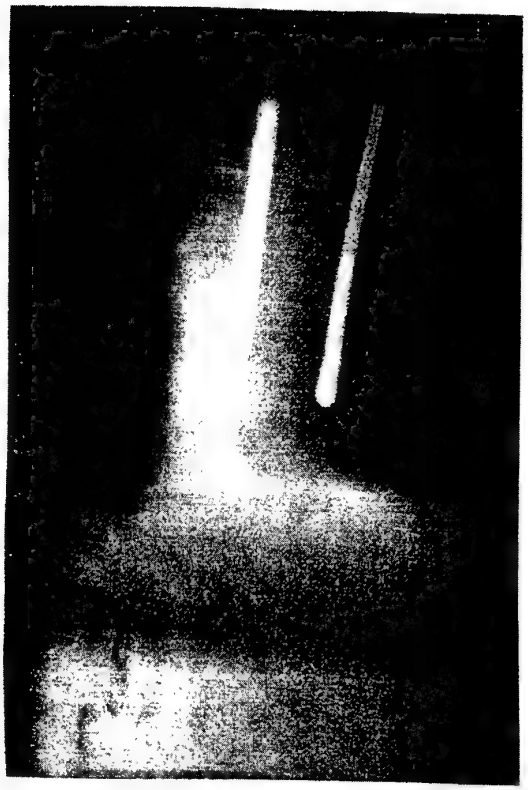
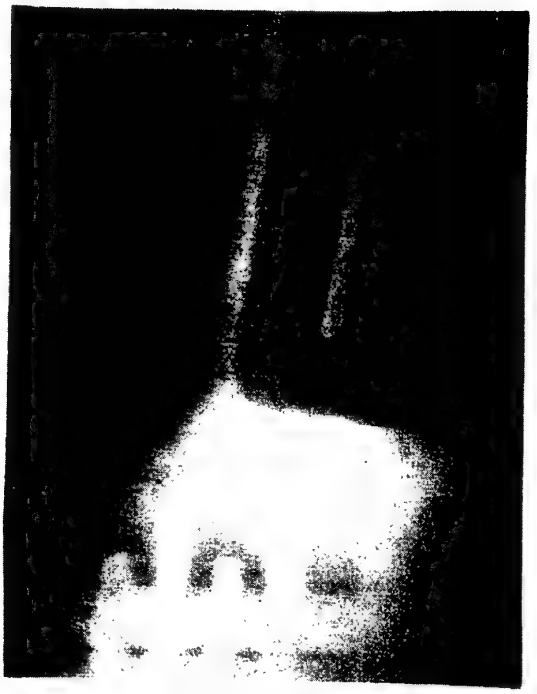


Fig. 11. Energy spectrum with and without filters of the blade area of the specimen.



(a) Without filter



(b) With 3 mm Cu-filter

Fig. 12. Radiographs of mottling indications on the turbine blade foot;
(a) Without filter
(b) With 3 mm Cu-filter

Fig. 13. Radiograph of the turbine blade scanned with double-slit system in horizontal direction, 155 kV, 1.0 mA, t: 155 min, D7 film, 0.02 Pb screens, FFD: 360 mm, a: 140 mm, specimen-slit distance: 50 mm, c: 56 mm, c_1: 1 mm, c_2: 4 mm, opening angle: 5.1°, scanning speed: 100 mm/min.

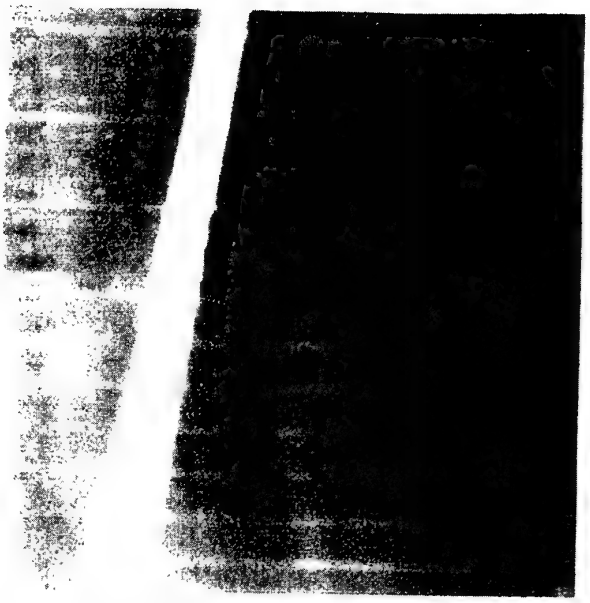


Fig. 15. Radiograph of a part of the blade with mottling indications, 160 kV, 1.0 mA, t: 12 min, D7 film, 0.02 Pb screens, FFD: 990 mm, a: 490 mm, focus-NDT slits distance: 700 mm.

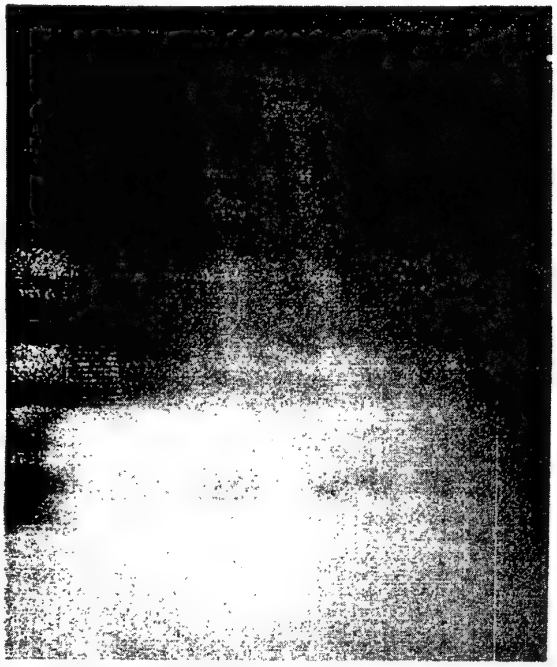


Fig. 14. Radiograph of the turbine blade scanned with double-slit system in vertical direction, 155 kV, 1.0 mA, t: 155 min, D7 film, 0.02 Pb screens, FFD: 360 mm, a: 140 mm, specimen-slit distance: 50 mm, c: 56 mm, c_1: 1 mm, c_2: 2 mm, opening angle: 3.27°, scanning speed: 100 mm/min.

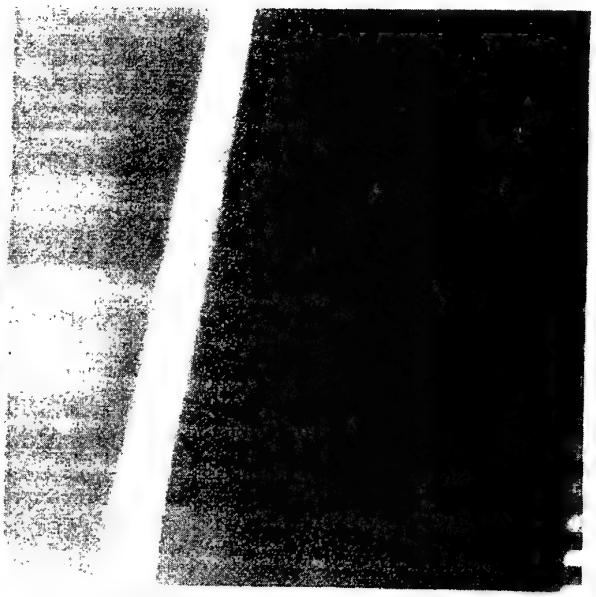


Fig. 16. Radiograph of the part of the blade with one vertical NDT-slit, 160 kV, 1.0 mA, t: 15 min, D7 film, 0.02 Pb screens, FFD: 990 mm, a: 490 mm, focus-NDT slits distance: 700 mm.

Fig. 17. Radiograph of the part of the blade with two NDT-slits in vertical and horizontal directions, 160 kV, 1.0 mA, t: 15 min, D7 film, 0.02 Pb screens, FFD: 990 mm, a: 490 mm, focus-NDT-slits distance: 700 mm.

BOREHOLE INSPECTION WITH ROTATING EC-PROBES
A NEW PROCEDURE WITH IMPROVED RELIABILITY

D. Schiller
H. Speckmann
Daimler-Benz Aerospace Airbus GmbH
Dept. EVP
28183 Bremen, Germany

SUMMARY

This report describes the development of a standard procedure for nondestructive inspection of cylindrical bores with rotating Eddy Current probes.
It can replace approximately 90% of the special inspection procedures included in the Nondestructive Testing Manual (NTM). It is independent of the type of aircraft and can be used on all aircraft structures which meet the requirements of the structure specification defined in this procedure.
One requirement to be met by the standard procedure was the verification of a probability of detection (POD) of 90% at a 95% confidence level for a fatigue crack with a maximum length ≥ 1mm. This was verified by means of a qualification program.

The basic development of the standard procedure was carried out on a European level in the framework of the BRITE/EURAM Program (BE5145) initiated by the European Community. The objective of this program was to increase the safety of aircraft by improving reliability, quality, and cost effectiveness of the inspection of safety-critical structures.

CONTENTS LIST

1. INTRODUCTION

Aircraft and engine manufacturers as well as airlines have their own Non-Destructive Testing (NDT) procedures with special characteristics of their own. This means that during aircraft maintenance, the NDT personnel constantly has to familiarize with different inspection procedures - despite similar inspection requirements.

This again not only leads to increased costs for equipment and adjustment standards, but also plays an important rule in the reproducibility and reliability of an inspection procedure.
The probability of a procedural error due to a large number of different and constantly changing inspection instructions is considerably higher as if there is one single procedure which is applied so often that it becomes "second nature".
The reliability of an inspection can be considerably increased by the use of one single procedure with standard characteristics, and costs for material and personnel are reduced at the same time.

Paper presented at the RTO AVT Workshop on "Airframe Inspection Reliability under Field/Depot Conditions", held in Brussels, Belgium, 13-14 May 1998, and published in RTO MP-10.

2. CURRENT SITUATION

Borehole inspection with rotating Eddy Current probes, the application of which celebrates its silver jubilee at the end of 1996, was selected as one of the most important inspection procedures. The reasons were as follows:

- No standard procedure available
- Potential of standardization
- The inspection method still includes questions, which have not been clarified up to now
- Some influencing factors still haven't been investigated

The various types of procedures for the inspection of boreholes with rotating Eddy Current probes mean that maintenance personnel constantly has to adjust to new conditions. In addition, a large number of different adjustment standards are used in the individual procedures.
For the user, these factors are very time-consuming and costly.
With a standard procedure, 90% of all borehole inspections required in the aerospace industry could be carried out, which would contribute to a considerable increase in reliability and effectivity.

3. PARAMETERS INVOLVED

The procedures for the inspection of boreholes with rotating Eddy Current probes sometimes contain considerable differences, which inevitably lead to a different evaluation of defects.
The hardware components that are likely to vary are shown in Figure 1.

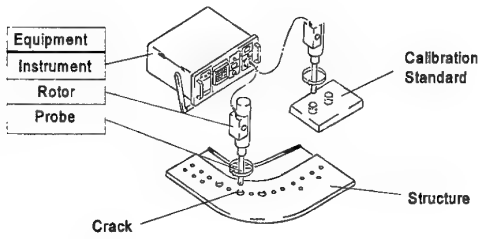


Figure 1 - Involved hardware

In detail the following factors may vary (selection):

- Equipment: Instrument — Solid Frequency; Variable Frequency
  - Rotor — Frequency (LF, MF,HF); Dimension
  - Probe — Coil-Core Diameter; Geometry
- Structure: Material; Conductivity; Single sheet; Multilayered; Thickness, single sheet; Thickness, complete structure; Combination of different materials; Cold expanded / not cold expanded
- Calibration Standard: Material; Geometry (cone, single sheet hole plate, multilayered hole plate); Slot geometry (corner slot, through slot); Kind of slot (eroded, saw cut, divided)
- Crack: Geometry; Depth / length; Position; Noise indications

4. DEVELOPMENT OF A STANDARD PROCEDURE

4.1. GENERAL

The standard procedure for the inspection of boreholes with rotating Eddy Current probes was developed for the inspection of typical aerospace structures.
The world wide most common rotating bolt hole inspection equipment is the Rototest (single frequency). Therefore this unit was selected for evaluation and the further development of the standard procedure.
The Rototest operates with a driver-receiver coil system and determines the frequency of 500 kHz for the standard rotor. The rotating speed lies between 2000 and 3000 RPM.
The rotating probes used in this procedure have the most common coil configuration: 1mm split core.

4.2. KEY ELEMENTS OF THE PROCEDURE

A standard procedure has been developed with the objective of combining the largest possible number of test procedures with similar requirements in a standard procedure which would facilitate handling. The wide application of such a procedure will considerably increase reliability.
The key elements are the following:

4.2.1. MATERIAL CONDUCTIVITY

The electrical and magnetic conductivity of an aircraft structure to be inspected plays an important role in Eddy Current techniques.
A standard procedure therefore has to specify on which materials an inspection may be carried out.
The standard procedure for borehole inspection with rotating Eddy Current probes may be used on materials with the following properties:

- Non-ferrous
- Conductivity range: 12 MS/m - 28 MS/m

Materials outside this conductivity range must not be inspected with the standard procedure, as this leads to deviations in the signal amplitude and to phase rotations.

4.2.2 SET-UP AND ACCESS

Instead of the detailed inspection area only applicable for one procedure that is shown in the special inspection procedures, the standard procedure includes a schematic set-up containing the minimum dimensions for accessibility determined by the equipment (refer to Figure 2).

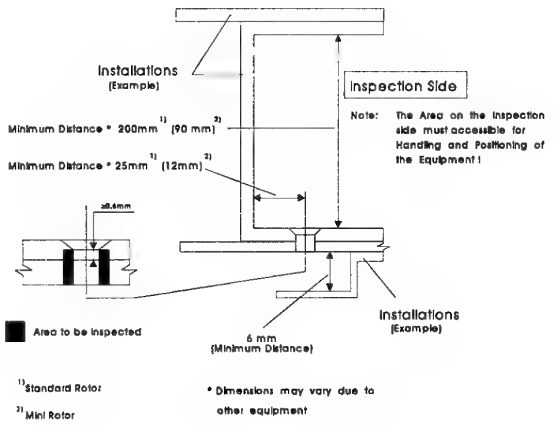


Figure 2 - Set-Up

4.2.3. DETECTABLE DAMAGE

In the framework of this development, the most frequently occurring cracks were determined and considered in the procedure (refer to Figure 3). In this context, the following factors influence the detectability of a crack:

- Crack geometry
- Crack position
- Thickness of layer with a crack

As in the determination of crack properties, the influence of the sheet thicknesses having cracks was also considered. Examinations on various structures (single layer, multi-layered) showed different defect amplitudes of cracks during the inspection of thin and thick sheets.
Therefore a differentiation in the crack geometry was made between cracks in sheets with a thickness between 0.6mm and 1.0mm, and cracks in sheets with a thickness 〉 1mm.

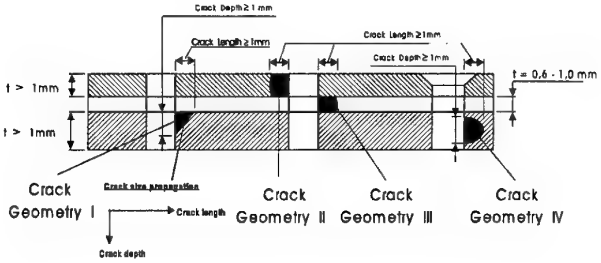


Figure 3 - Detectable crack types

4.2.4. CALIBRATION STANDARD

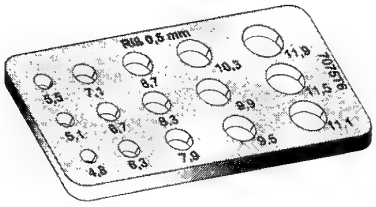
The editors evaluated the properties of the most common standards, characterized their properties and compared the sensitivity and the correction factor. The most important requirement was the true phase indication. This means that the eddy current signal produced by the calibration standard must have the same phase angle as the real cracks. This feature is essential for the phase out technique. Further on the handling is one important item. It was found that the set-up of the sensitivity amplitude of a signal was more difficult with a fixed diameter hole standard. Reason for the obtained variation was the possible air gap between probe and hole. Additionally wear and manufacturing tolerance amplify this deviation. And a very important point for the airlines, standards must be cheap and for universal use.

Evaluation of existing standards

Type 1: Plate
Common standard with hole diameters equivalent to the aircraft manufacturers requirement. The defects are simulated by saw cuts or EDM notches.

The defect size:	Width	0,15 mm
	Depth	0,5 mm
	Length	plate thickness.

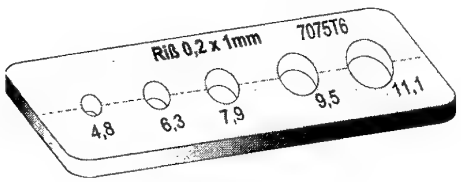
Due to the air gap between probe and standard it can be difficult to maintain the same sensitivity setting for recalibration.



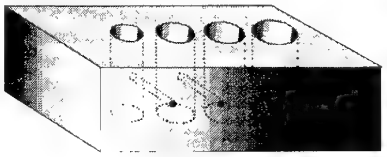
Type 2: Plate with surface slot
A similar standard as type 1 but with a different defect simulation with a good
width / depth ratio and a true type phase indication.

The defect size:	Width	0,2 mm
	Depth	1,0 mm
	Length	plate length.

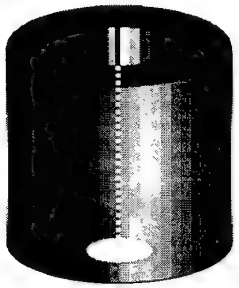
Due to the air gap between probe and standard it can be difficult to maintain the same sensitivity setting for recalibration. Additionally there are two crack indications on the CRT display which will certainly have a different amplitude. For calibration the operator has to focus on one defect and try to rise this signal to the maximum.



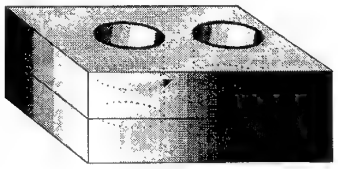
Type 3: Plate with side holes
This standard has hole diameters as required with a hole of 1,6 mm drilled from the side. The phase indication is rotated. Filter setting is difficult due to the open loop signal which will not indicate as a symmetric shape. But the part is easy to manufacture and cheap and may be "home made".



Type 4: Conical Standard / Förster
The basic idea of the conical standard was to cover a wide range of fastener hole diameters. The defect is simulated by a 0,3 mm deep and 0,15 mm wide EDM notch. The ratio D / W of 2/1 is very poor and this results in a significant phase rotation. Damage and wear may also cause variation in sensitivity. A rework of the standard is possible but expensive.



Type 5: Multi-layer plate
This standard simulates a typical aircraft structure, including the defect type and location which reflects the characteristics of possible cracks. The defect ranges from 0,2 mm up to 1 mm, all approx. 0,15 mm wide and is manufactured as saw cut or EDM notch. Due to the 45° slot there is a variation of the depth to width ratio which ranges from 1/1 up to 3,3/1. But again the disadvantage of this unit is the non-realistic phase indication.



Type 6: Conical standard / Brite Euram type
This standard looks very much like the Förster cone. But the basic idea is completely different. Two halves of a rod are fitted together and machined to the final shape. The tiny gap between the two layers simulates the defect. There are no expensive EDM notches and the ratio of crack width to length is much better than 1/10. There are no depth variations or changes of the defect and therefore no variations in sensitivity setup. Rework of the BE cone is easily possible.

Summary:

As a result of this evaluation we selected the newly developed Splitted- Conical-Calibration-Standard (SC^2S) (refer to Figure 4) for some trial inspections and the comparison of the new procedure with the existing ones. One remarkable result was that simulated defects with a depth of 1 mm or more did not produce any larger signals than 0,5 mm deep defects.

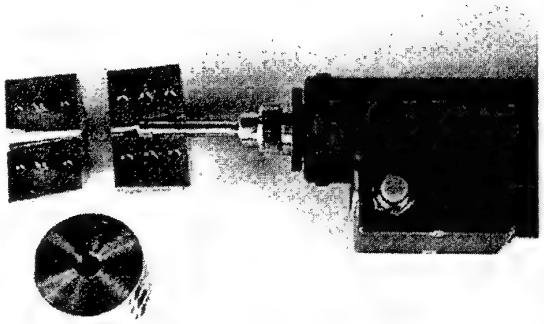


Figure 4 - Splitted Conical Calibration Standard (SC^2S) BRITE/EURAM TYPE

4.2.5. INSTRUMENT CALIBRATION WITH THE SENSITIVITY CORRECTION VALUE

Based on the most frequent types of cracks and the most common adjustment standards, a correction value is added to the basic adjustment (100% SH above zero datum), which compensates the differences (refer to Table 1 and Figure 3), so that the inspection sensitivity is always the same for a certain type of defect and different standards.

Crack-Geometry to be detected	Additional corrections [ΔdB] for equipment adjustment with			
	Cone with Split-Plane (Ref.Fig.4)		Cone with slot depth 0,3mm (e.g. Förster-Cone)	Cone or plate with slot depth 0,5mm
I	+ 10		+ 6	+ 10
II	0		- 4	0
III	thickn. = 0,6mm	+ 10	+ 6	+ 10
	thickn. = 0,8mm	+ 6	+ 2	+ 6
	thickn. = 1,0mm	+ 2	- 2	+ 2
IV	+ 10		+ 6	+ 10

Table 1- Correction values

4.2.6. EVALUATION OF INDICATION

For signal evaluation and interpretation the Rototest equipped with both, the Y-T and X-Y CRT mode, and only the use of both provides the operator with the full information.
But most operators still do not use the X-Y display capability of their instruments for evaluation of the signals.

The Y-T mode is ideal for the evaluation of the number of cracks and their location in the bolt hole.
Additionally the amplitude of the signal is clearly visible and the shape of the display contains some limited phase information.
To gain the full phase information, the X-Y mode is essential to set-up the instrument in a proper way.
Cracks and other defects like damages, pilot holes, foreign material or shims have their own unique "fingerprint" display.
Two significant examples are shown in the standard procedure and in Figure 5.

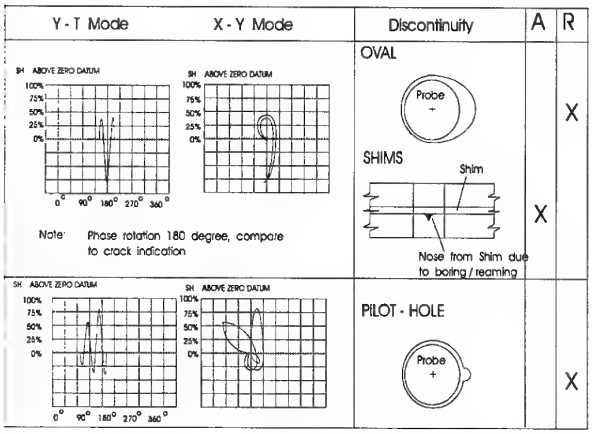


Figure 5 - Signals due to discontinuities

4.2.7. APPLICATION OF PHASE-OUT TECHNIQUE

The condition of fastener holes and the poor quality of workmanship may cause misreadings during an eddy current inspection. These indications, called noise, are created by foreign metal deposits like cadmium burr caused by fastener removal or reaming mechanical damages, ovality and eccentricity and corrosion.
Economic and safety reasons require the reduction of this kind of possible misinterpretation of indications.
One of the most efficient procedures to improve the quality of inspection and reduce the number of false calls is the so called phaseout technique. The basic principle of this technique is the elimination of the non-relevant signals by a phase rotation.

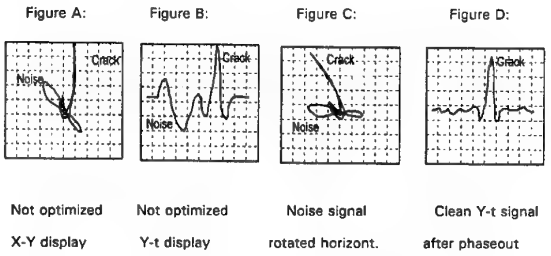


Fig. A shows a crack signal pointing in upscale direction and a possible noise indication at 45° position of the screen. **Fig. B** shows the resulting Y-t display normally used for inspection. Both noise and crack will indicate upscale. Even the evidence of the noise would cause a rejection of the hole, followed by further action like polishing or oversizing and reinspection. The use of the phaseout technique as demonstrated in **Fig. C** will eliminate the indications caused by the noise and only real cracks will be found during the inspection as shown in **Fig. D**.
During the evaluation of the phaseout technique we discovered one significant problem. The common calibration standards with artificial cracks simulated by saw cuts or EDM notches caused a phase deviation of up to 30° in counter clockwise direction.
As a result of this peculiarity there was no sufficient phase separation between noise and expected crack phase indication. The technique was not usable. This phenomenon caused a further evaluation of the common calibration standards. The editors looked at several, frequently used standards tried to find a "true phase" calibration standard and finally developed a new one.

4.2.8. PROBABILITY OF DETECTION (POD)

In the framework of the development of this inspection procedure, a qualification (refer to chapter 6.) was carried out with the objective of verifying the reliability of the detection of the a.m. types of cracks.
A reliability of 90% with a confidence level of 95% was to be demonstrated (refer to figure 6).

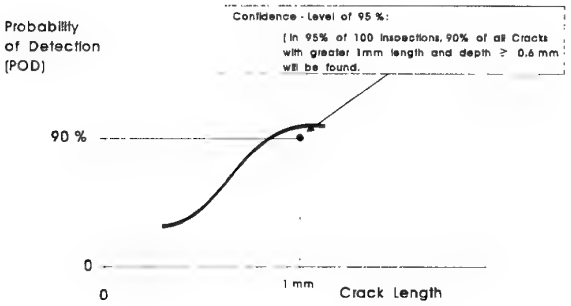


Figure 6 - POD curve

5. AIRLINE VISITS

5.1. GENERAL

After development of the new Standard Procedure we planned to run some field tests with experienced NDT inspectors from international airlines.

Reasons for visits were:

- Evaluate the "state of the art" in bolt hole inspection
- Get comments on manufacturer procedures
- Compare: Airline procedure - New Standard Procedure
- Collect comments and improvements
- Collect reliability data
- Have the procedure approved by NDT specialists

5.2. HARDWARE USED FOR TEST RUNS

Test specimen
For the test runs at the airline facilities a specific test piece was designed. Basically a cut-out of an old aircraft frame was used, containing approx. 30 holes of 1/4" diameter (See Fig. 7).

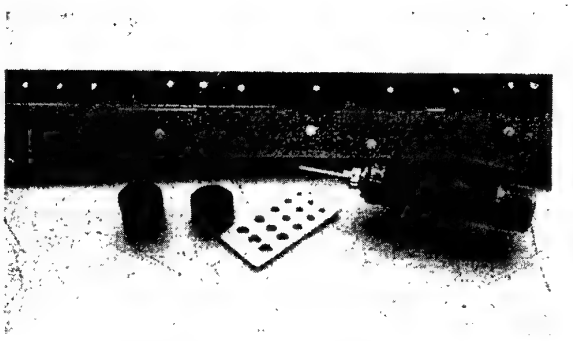


Figure 7 - Hardware for airline visit

To cover the variety of inspection conditions this part was widely modified. Up to 4 layers of sheet metal were added, containing fatigue cracks with different lengths. Additionally other defects like damage, pilot holes, ovality, splices were included to simulate the field conditions and make the inspection more difficult.

Calibration standard
In addition to the airlines own standard the new developed conical standard, was used in the second test run.

Eddy Current equipment
For all trials the Rototest was used, including a standard rotor and probes.
For those operators who did not know the unit, a short familiarisation training was conducted.

5.3 PERFORMANCE OF ROTOTEST BOLT HOLE INSPECTION TEST RUN

After a general introduction, the reason for the visit was explained in detail. The inspector was told that it was not the goal of the test run to evaluate his performance. Instead it was actually planned to checkout the newly developed standard procedure in comparison to the airlines daily routine procedure. The reporting system and the documentation sheets were shown and explained. The test was divided into two parts.
The first run was conducted with the airlines own procedure and their common calibration standard and setup. This test simulated the actual state of the art.
After completion, the second run was arranged using the new procedure and the conical standard.

5.4. RESULTS OF TRIALS

5.4.1. SUMMARY OF EVALUATION

A lot of interesting data were collected during the visits. Instead of only preparing statistical data, we tried to extract and summarize the important information. This information is shown in the following tables 2 and 3:

Table 2: Airline own procedure
Table 3: New Standard Procedure

Airlines own Procedure									
Airline / Operator	A1	A2	B1	B2	C1	D1	E1	F1	F2
Written Procedure avail.	no	no	no	no	no	no	no	no	no
Language of Procedures	own	own	E	E	E	E	E	E	E
Type of Cal.-Std. used	1	1	2	2	3	I	1	equiv. to 5	
Defect type	EDM	EDM	Saw	Saw	Hole	Saw	Saw	EDM	EDM
Defect size (mm)	0.5	0.5	>>	>>	>>	0.5	0 5	unknown*1	
Set to Amplitude %	100	100	75	75	40 +10dB	100	100	140	100
Handling of equipment	+	+	+	-	+	+	+	+	-
Setup of Rototest	ok	ok	ok	ok	ok	ok	ok	ok	ok
Filter setting	ok	ok	ok	ok	diffic.	ok	ok*2	ok	ok
Phase setting to	top	top	top	top	30°left	top	top	top	top
Applies visual first	no	no	no	no	no	no	no	no	no
Performance of insp.	+	+	+ +	+-	+ +	slow		slow	ok
Experienced Inspector	no	no	yes	+-	yes	+-	yes	+-	yes
Inspection mode Y-t	se-arch	se-arch	se-arch	se-arch	search	se-arch	locate	se-arch	search
Inspection mode X-Y	no	no	no	no	search	no	eval.	no	eval.
Familiar w. x-y evaluation	no	no	no	no	no	no	yes	no	yes
Uses phase evaluation	no	no	no	no	yes	no	yes	no	yes
Evaluates signalshape	no	no	yes	yes	yes	yes	yes	no	yes

Remarks:

*1) No details available for cal. std. Manufacturer and crack depth unknown.

*2) Inspector changed filter setting after first hole.

*3) Did a visual inspection after eddy current testing.

Table 2 - Airline Procedure

BE 5145 Procedure									
Airline / Operator	A1	A2	B1	B2	C1	D1	E1	F1	F2
Speaks English	+-	+-	yes	yes	yes	yes	yes	yes	yes
Reads BE Procedure	transl	transl	yes	no* 1)	yes	yes	yes	yes	yes
Understands BE Proc	yes	yes	yes	no	yes	yes	yes	yes	yes
Asks for details	no	no	yes	no	yes	yes	no	no	yes
Uses Conical correct	yes	yes	yes	yes	yes	yes	yes	yes	yes
Applies dB Correction	yes	yes	yes	yes	yes	yes	yes	yes	yes
Sets Phase correct	yes	yes	yes	yes	yes	yes	yes	yes	45-°left
Handling of equipment	good	good	good	+-	good	good	good	good	ok
Applies visual first	no	no	yes	no	yes	yes	no	yes	yes
Performance of inspection	ok	ok	good	+-	good	ok	ok	ok	ok
Inspection mode Y-t	search	search	search	search	search	search	search	search	search
Inspect mode X-Y	eval	eval	eval	-	eval	-	eval	eval	eval
Uses phase evaluation	yes	yes	yes	no	yes	no	yes	+-	yes
Uses phaseout	yes	yes	no	no	yes	yes	no	no	yes
dB corr. after phase-out	no	no	-	-	no	no	-	-	no
Evaluates signal shape	no	no	yes	no	yes	yes	yes	+-	yes

Remarks:

*1) Inspector did not like to read long procedures. He found it easier to follow the pictures.

Table 3 - New Standard Procedure

5.4.2. REPORTED FINDINGS

The inspected part contained 5 cracked bolt holes and 6 other indications which should be reported by the operator.
For the comparison of the two trials with the airline procedure and the new procedure the matching data are shown in the graph below.
The reference bar indicates the best possible result in terms of hits and misses.

It is clearly visible that the new procedure results in less findings which means less false calls due to noise and others.
The two lower graphs in the display show the number of missed real cracks.
Again the new procedure missed 2 cracks and the airlines procedure missed 5 cracks.

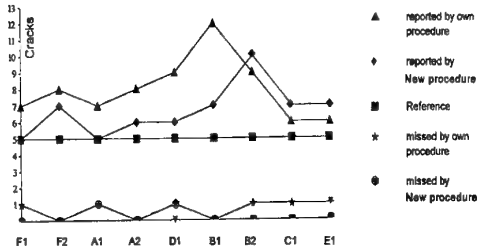


Diagram 1 - Reported and missed cracks

Looking in detail at the overall performance, only the inspectors M2, A2 and B1 did not miss any cracks. But in case of B1, the inspector recorded much more false calls than others, amounting to a total of 12 crack indications which means he had 7 not cracked holes rejected.
But anyway the data must be seen in connection with the reported other damages as shown in the following graph.

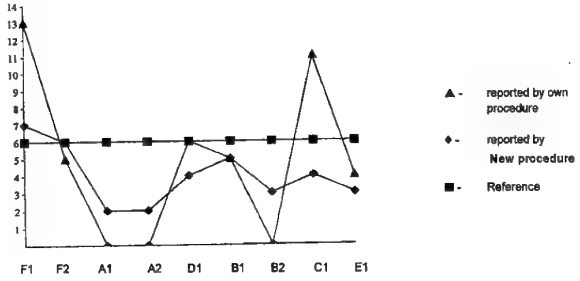


Diagram 2 - Reported non-crack damages

Again one can see the same tendency as before. Using the new procedure the variations are much smaller. The deviation from the reference value is at the maximum +1 and at the minimum -4 indications, which is a remarkable improvement of the inspection.
As a final conclusion it can be stated that the improvements of inspection results are remarkable.
The newly developed procedure as well as the new calibration standard may help airlines to reduce labour cost and repair time by a significant amount.
In addition, it is an important improvement of inspection quality and safety.

5.4.3 COMMENTS OF INSPECTORS

In a final debriefing the inspectors could comment the new standard procedure and the newly developed conical calibration standard.

Positive

Sample indications very good
New conical st. also wanted for steel and titanium
Got new information on phase evaluation
Conical very useful
POD good information
Procedure is a dramatic step forward
Cone is easy to use, more user-friendly, cheaper than others, true phase indication
It was a good free of charge training lesson
Gain correction was a new information
Phaseout not applied before

Negative

Procedure should be better structured
Flow chart may be useful

6. QUALIFICATION PROGRAM

6.1. GENERAL

The qualification program was carried out with the following objectives:

- Verification of the crack geometry required in the inspection procedure (refer to Figure 9) with a probability of detection (POD) of 90% and a confidence level of 95%.

- Determination of the false alarm rate (target ≤3,0%).

6.2. QUALIFICATION PROCEDURE

- For the qualification program the following specimen was used:

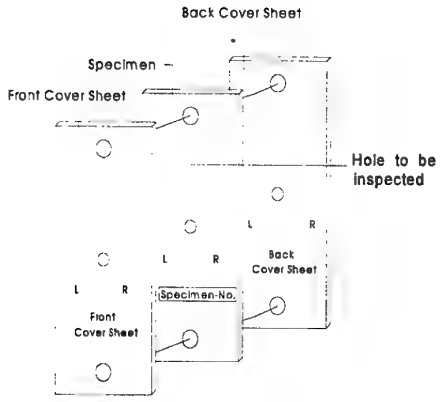


Figure 8 - Specimen type

This specimen batch consists of 119 plates with two different hole diameters (4.9 and 6.4 mm) and sheet thicknesses of 0.6mm, 1.6mm and 3.0mm (refer to table 4).
The holes in these specimens are very smooth, because there were no rivets inside which had to be removed before inspection. This means that the batch is similar to inspections of clean reamed and polished fastener holes. The cracks to be detected are natural fatigue cracks and in the range of 0mm to 5mm in length. The holes to be inspected could be affected with one or two cracks.

Hole dia. [mm]	Sheet thickn [mm]	Number of plates	Number of cracks[1]	Number of inspectors	Specimen: Areas with cracks[2]	Specimen: Total areas to be inspected[3]
4,9	0,6	14	14	7	98	196
	1,6	13	12		84	182
	3,0	15	17		119	210
6,4	0,6	31	33		231	434
	1,6	19	17		119	26
	3,0	27	30		210	378
TOTAL		119	123	TOTAL	861	1666

[1] Cracks ≤0,35 mm are not considered in the qualification

[2] Number of cracks x 7 (Number of inspectors)

[3] Number of plates x 7 (Number of inspectors) x 2

Table 4 - Qualification program

- The specimens were examined by 7 inspectors with level 1 or 2 certification:
  - 2 inspectors from OGMA, Portugal
  - 1 inspector form SAAB, Sweden
  - 1 inspector from Airbus Industrie, France
  - 1 inspector from British Aerospace Airbus, UK
  - 1 inspector from Alitalia, Italy
  - 1 inspector from Deutsche Lufthansa, Germany

- The qualification was carried out in accordance with the new standard procedure:

  ROTATING PROBE INSPECTION OF FASTENER HOLES WITH EDDY CURRENT (1)

and the new Split Conical Calibration Standard (SC^2S).

The inspectors had to choose the appropriate correction value for the defined crack type (refer to Figure 9). In this case a correction value of +10 dB had to be used.

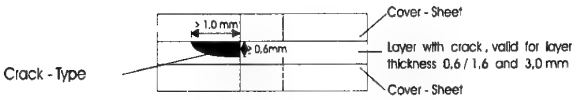


Figure 9 - Crack type to be found

6.3 DATA EVALUATION AND RESULTS

An earlier investigation, made at Daimler-Benz Aerospace Airbus, showed that the influence of sheet thicknesses for crack detection can be divided into two groups:

GROUP 1: ≤ 1.0mm (in the qualification: 0.6mm sheet thickness)
GROUP 2: > 1.0mm (in the qualification: 1.6 and 3.0mm sheet thickness)

For the evaluation, only cracks ≥0.35 mm are considered.

■ 0.6mm sheet thickness

Diagram 3 shows the results of the evaluation of the data from all inspectors.
The essential information is that with a POD of 90% and a confidence level of 95% a crack ≥0.5mm can be found .
The false alarm rate for sheets of 0.6mm thickness is 1.5%.

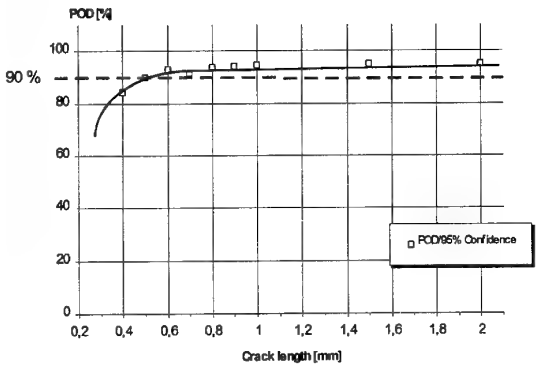


Diagram 3 - POD curve for 0.6mm sheet thickness

■ 1,6mm / 3,0mm sheet thickness

The curves in the diagram 4 for sheets of 1,6mm and 3,0mm thickness are similar to the diagram 3. The essential information is, that with a POD of 90% and a confidence level of 95% a crack ≥0,7mm can be found .

The false alarm rate for sheets of 1,6mm and 3,0mm thickness is 2,5%.

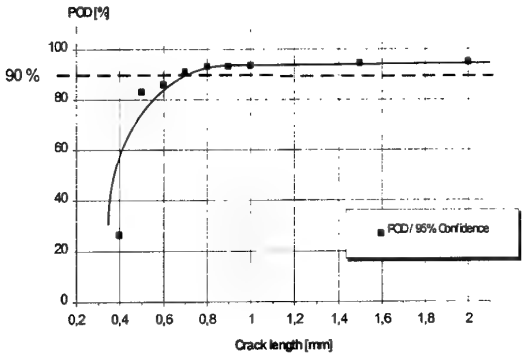


Diagram 4 - POD curve of 1,6mm and 3,0mm sheet thickness

7. CONCLUSION

The new inspection procedure with its standard procedure characteristics very well fulfills and even exceeds the current inspection requirements.
Therefore, the new Splitted-Conical-Calibration-Standard applied here has met the requirements in comparison with previously used calibration standards.
With this result and the totally positive airline statements, this inspection procedure furnishs proof of a clearly measurable improvement in bore inspection reliability.
In short terms it could be said, that the benefits of the developed standard procedure for "Borehole inspection with rotating EC-Probes" are:

- Improved the reliability of airframe inspection
- Short response to new test problems, due to short preparation time
- Improved the effectiveness of the inspection technique
- No extra qualification/validation in case of standard procedure application
- In approx. 90% of all bore hole inspections, the standard procedure can be used
- Reduction of calibration and handling errors
- Reduced material costs
- If SC^2S is used, only one adjustment standard is needed
- Existing standards can still be used applying the sensitivity correction
- By applying the sensitivity correction value the sensitivity can be optimized to suit the type of crack to be detected
- Improved interpretation of test results due to Phase-Out technique
- Suppression of disturbance signals e.g. shims, pilot hole, etc.

8. REFERENCES

(1) BRITE/EURAM Report BE 145R030

REPORT DOCUMENTATION PAGE

1. Recipient's Reference	2. Originator's References	3. Further Reference	4. Security Classification of Document
	RTO MP-10 AC/323(AVT)TP/2	ISBN 92-837-1002-9	UNCLASSIFIED/ UNLIMITED

5. Originator Research and Technology Organization
North Atlantic Treaty Organization
BP 25, 7 rue Ancelle, F-92201 Neuilly-sur-Seine Cedex, France

6. Title
Airframe Inspection Reliability under Field/Depot Conditions

7. Presented at/sponsored by
The Workshop of the RTA Applied Vehicle Technology Panel (organised by the former AGARD Structures and Materials Panel) held at Quartier Reine Elisabeth, Etat-Major de la Force Aérienne, Brussels, Belgium, 13-14 May 1998

8. Author(s)/Editor(s)	9. Date
Multiple	November 1998
10. Author's/Editor's Address	**11. Pages**
Multiple	202

12. Distribution Statement There are no restrictions on the distribution of this document. Information about the availability of this and other RTO unclassified publications is given on the back cover.

13. Keywords/Descriptors

Nondestructive tests	Aircraft maintenance
Airframes	POD (Probability of Detection)
Airworthiness	Damage assessment
Fatigue (materials)	Inspection
Service life	Reliability
Aircraft	Field tests
Failure analysis	Quality assurance

14. Abstract

Contains the papers presented at a Workshop on Airframe Inspection Reliability under Field/Depot Conditions organised by the Applied Vehicle Technology Panel (AVT) of RTO, in Brussels, Belgium, 13-14 May 1998.

The Workshop had the general objective of promoting general discussion on the merits and practicality of generating NDI Probability of Detection (POD) from in-service data and on the use of reliability data in the life-cycle management process.

The papers are presented under the following headings:

- Perspectives on: (i) the role of NDI, (ii) factors influencing eddy current POD in the field environment, and (iii) NDT reliability
- Estimation from small samples and in-service experience
- Approaches to POD generation
- Analytical issues related to generation and use of POD data
- Practical experience and case studies

NORTH ATLANTIC TREATY ORGANIZATION

RESEARCH AND TECHNOLOGY ORGANIZATION
BP 25 • 7 RUE ANCELLE
F-92201 NEUILLY-SUR-SEINE CEDEX • FRANCE
Télécopie 0(1)55.61.22.99 • Télex 610 176

DIFFUSION DES PUBLICATIONS
RTO NON CLASSIFIEES

L'Organisation pour la recherche et la technologie de l'OTAN (RTO), détient un stock limité de certaines de ses publications récentes, ainsi que de celles de l'ancien AGARD (Groupe consultatif pour la recherche et les réalisations aérospatiales de l'OTAN). Celles-ci pourront éventuellement être obtenues sous forme de copie papier. Pour de plus amples renseignements concernant l'achat de ces ouvrages, adressez-vous par lettre ou par télécopie à l'adresse indiquée ci-dessus. Veuillez ne pas téléphoner.

Des exemplaires supplémentaires peuvent parfois être obtenus auprès des centres nationaux de distribution indiqués ci-dessous. Si vous souhaitez recevoir toutes les publications de la RTO, ou simplement celles qui concernent certains Panels, vous pouvez demander d'être inclus sur la liste d'envoi de l'un de ces centres.

Les publications de la RTO et de l'AGARD sont en vente auprès des agences de vente indiquées ci-dessous, sous forme de photocopie ou de microfiche. Certains originaux peuvent également être obtenus auprès de CASI.

CENTRES DE DIFFUSION NATIONAUX

ALLEMAGNE
Fachinformationszentrum Karlsruhe
D-76344 Eggenstein-Leopoldshafen 2

BELGIQUE
Coordinateur RTO - VSL/RTO
Etat-Major de la Force Aérienne
Quartier Reine Elisabeth
Rue d'Evere, B-1140 Bruxelles

CANADA
Directeur - Gestion de l'information
(Recherche et développement) - DRDGI 3
Ministère de la Défense nationale
Ottawa, Ontario K1A 0K2

DANEMARK
Danish Defence Research Establishment
Ryvangs Allé 1
P.O. Box 2715
DK-2100 Copenhagen Ø

ESPAGNE
INTA (RTO/AGARD Publications)
Carretera de Torrejón a Ajalvir, Pk.4
28850 Torrejón de Ardoz - Madrid

ETATS-UNIS
NASA Center for AeroSpace Information (CASI)
Parkway Center, 7121 Standard Drive
Hanover, MD 21076

FRANCE
O.N.E.R.A. (Direction)
29, Avenue de la Division Leclerc
92322 Châtillon Cedex

GRECE
Hellenic Air Force
Air War College
Scientific and Technical Library
Dekelia Air Force Base
Dekelia, Athens TGA 1010

ISLANDE
Director of Aviation
c/o Flugrad
Reykjavik

ITALIE
Aeronautica Militare
Ufficio Stralcio RTO/AGARD
Aeroporto Pratica di Mare
00040 Pomezia (Roma)

LUXEMBOURG
Voir Belgique

NORVEGE
Norwegian Defence Research Establishment
Attn: Biblioteket
P.O. Box 25
N-2007 Kjeller

PAYS-BAS
RTO Coordination Office
National Aerospace Laboratory NLR
P.O. Box 90502
1006 BM Amsterdam

PORTUGAL
Estado Maior da Força Aérea
SDFA - Centro de Documentação
Alfragide
P-2720 Amadora

ROYAUME-UNI
Defence Research Information Centre
Kentigern House
65 Brown Street
Glasgow G2 8EX

TURQUIE
Millî Savunma Başkanliği (MSB)
ARGE Dairesi Başkanliği (MSB)
06650 Bakanliklar - Ankara

AGENCES DE VENTE

NASA Center for AeroSpace Information (CASI)
Parkway Center
7121 Standard Drive
Hanover, MD 21076
Etats-Unis

The British Library Document Supply Centre
Boston Spa, Wetherby
West Yorkshire LS23 7BQ
Royaume-Uni

Canada Institute for Scientific and Technical Information (CISTI)
National Research Council
Document Delivery,
Montreal Road, Building M-55
Ottawa K1A 0S2
Canada

Les demandes de documents RTO ou AGARD doivent comporter la dénomination "RTO" ou "AGARD" selon le cas, suivie du numéro de série (par exemple AGARD-AG-315). Des informations analogues, telles que le titre et la date de publication sont souhaitables. Des références bibliographiques complètes ainsi que des résumés des publications RTO et AGARD figurent dans les journaux suivants:

Scientific and Technical Aerospace Reports (STAR)
STAR peut être consulté en ligne au localisateur de ressources uniformes (URL) suivant:
http://www.sti.nasa.gov/Pubs/star/Star.html
STAR est édité par CASI dans le cadre du programme NASA d'information scientifique et technique (STI)
STI Program Office, MS 157A
NASA Langley Research Center
Hampton, Virginia 23681-0001
Etats-Unis

Government Reports Announcements & Index (GRA&I)
publié par le National Technical Information Service
Springfield
Virginia 2216
Etats-Unis
(accessible également en mode interactif dans la base de données bibliographiques en ligne du NTIS, et sur CD-ROM)

Imprimé par le Groupe Communication Canada Inc.
(membre de la Corporation St-Joseph)
45, boul. Sacré-Cœur, Hull (Québec), Canada K1A 0S7

NORTH ATLANTIC TREATY ORGANIZATION

RESEARCH AND TECHNOLOGY ORGANIZATION
BP 25 • 7 RUE ANCELLE
F-92201 NEUILLY-SUR-SEINE CEDEX • FRANCE
Telefax 0(1)55.61.22.99 • Telex 610 176

DISTRIBUTION OF UNCLASSIFIED RTO PUBLICATIONS

NATO's Research and Technology Organization (RTO) holds limited quantities of some of its recent publications and those of the former AGARD (Advisory Group for Aerospace Research & Development of NATO), and these may be available for purchase in hard copy form. For more information, write or send a telefax to the address given above. **Please do not telephone**.

Further copies are sometimes available from the National Distribution Centres listed below. If you wish to receive all RTO publications, or just those relating to one or more specific RTO Panels, they may be willing to include you (or your organisation) in their distribution.

RTO and AGARD publications may be purchased from the Sales Agencies listed below, in photocopy or microfiche form. Original copies of some publications may be available from CASI.

NATIONAL DISTRIBUTION CENTRES

BELGIUM
Coordinateur RTO - VSL/RTO
Etat-Major de la Force Aérienne
Quartier Reine Elisabeth
Rue d'Evere, B-1140 Bruxelles

CANADA
Director Research & Development
Information Management - DRDIM 3
Dept of National Defence
Ottawa, Ontario K1A 0K2

DENMARK
Danish Defence Research Establishment
Ryvangs Allé 1
P.O. Box 2715
DK-2100 Copenhagen Ø

FRANCE
O.N.E.R.A. (Direction)
29 Avenue de la Division Leclerc
92322 Châtillon Cedex

GERMANY
Fachinformationszentrum Karlsruhe
D-76344 Eggenstein-Leopoldshafen 2

GREECE
Hellenic Air Force
Air War College
Scientific and Technical Library
Dekelia Air Force Base
Dekelia, Athens TGA 1010

ICELAND
Director of Aviation
c/o Flugrad
Reykjavik

ITALY
Aeronautica Militare
Ufficio Stralcio RTO/AGARD
Aeroporto Pratica di Mare
00040 Pomezia (Roma)

LUXEMBOURG
See Belgium

NETHERLANDS
RTO Coordination Office
National Aerospace Laboratory, NLR
P.O. Box 90502
1006 BM Amsterdam

NORWAY
Norwegian Defence Research Establishment
Attn: Biblioteket
P.O. Box 25
N-2007 Kjeller

PORTUGAL
Estado Maior da Força Aérea
SDFA - Centro de Documentação
Alfragide
P-2720 Amadora

SPAIN
INTA (RTO/AGARD Publications)
Carretera de Torrejón a Ajalvir, Pk.4
28850 Torrejón de Ardoz - Madrid

TURKEY
Millî Savunma Başkanliği (MSB)
ARGE Dairesi Başkanliği (MSB)
06650 Bakanliklar - Ankara

UNITED KINGDOM
Defence Research Information Centre
Kentigern House
65 Brown Street
Glasgow G2 8EX

UNITED STATES
NASA Center for AeroSpace Information (CASI)
Parkway Center, 7121 Standard Drive
Hanover, MD 21076

SALES AGENCIES

NASA Center for AeroSpace Information (CASI)
Parkway Center
7121 Standard Drive
Hanover, MD 21076
United States

The British Library Document Supply Centre
Boston Spa, Wetherby
West Yorkshire LS23 7BQ
United Kingdom

Canada Institute for Scientific and Technical Information (CISTI)
National Research Council
Document Delivery,
Montreal Road, Building M-55
Ottawa K1A 0S2
Canada

Requests for RTO or AGARD documents should include the word 'RTO' or 'AGARD', as appropriate, followed by the serial number (for example AGARD-AG-315). Collateral information such as title and publication date is desirable. Full bibliographical references and abstracts of RTO and AGARD publications are given in the following journals:

Scientific and Technical Aerospace Reports (STAR)
STAR is available on-line at the following uniform resource locator:
http://www.sti.nasa.gov/Pubs/star/Star.html
STAR is published by CASI for the NASA Scientific and Technical Information (STI) Program
STI Program Office, MS 157A
NASA Langley Research Center
Hampton, Virginia 23681-0001
United States

Government Reports Announcements & Index (GRA&I)
published by the National Technical Information Service
Springfield
Virginia 22161
United States
(also available online in the NTIS Bibliographic Database or on CD-ROM)

Printed by Canada Communication Group Inc.
(A St. Joseph Corporation Company)
45 Sacré-Cœur Blvd., Hull (Québec), Canada K1A 0S7

ISBN 92-837-1002-9